



SIGUENOS EN:



LIBROS UNIVERISTARIOS Y SOLUCIONARIOS DE
MUCHOS DE ESTOS LIBROS GRATIS EN
DESCARGA DIRECTA

VISITANOS PARA DESARGALOS GRATIS.

Cálculo Diferencial de Varias Variables

Carlos Fernández Pérez
Francisco José Vázquez Hernández
José Manuel Vegas Montaner

Cálculo

Diferencial

de Varias Variables

Cálculo Diferencial de Varias Variables

Carlos Fernández Pérez
Facultad de Ciencias Matemáticas
Universidad Complutense de Madrid

Francisco José Vázquez Hernández
Facultad de Ciencias Económicas y Empresariales
Universidad Autónoma de Madrid

José Manuel Vegas Montaner
Facultad de Ciencias Matemáticas
Universidad Complutense de Madrid

THOMSON



Australia • Canadá • México • Singapur • España • Reino Unido • Estados Unidos

THOMSON

Cálculo diferencial de varias variables

© Carlos Fernández, Francisco José Vázquez y José Manuel Vegas

Gerente Editorial Área Universitaria:
Andrés Otero Reguera

Editoras de Producción:
Clara M^a de la Fuente Rojo
Consuelo García Asensio

Producción Industrial:
Susana Pavón Sánchez

Diseño de cubierta:



Impresión:
Closas Orocoyen, S.L.
Polígono Igarsa
nave 21, 22, 23 y 24
Paracuellos de Jarama
(Madrid)

**COPYRIGHT © 2002 International
Thomson Editores Spain
Paraninfo, S.A.
1ª edición, 2ª reimpresión, 2003**

Magallanes, 25; 28015 Madrid
ESPAÑA
Teléfono: 91 4463350
Fax: 91 4456218
clientes@paraninfo.es
www.paraninfo.es

Impreso en España
Printed in Spain

ISBN: 84-9732-056-5
Depósito Legal: M-31.623-2003

(072/71/54)

Reservados los derechos para todos los países de lengua española. De conformidad con lo dispuesto en el artículo 270 del Código Penal vigente, podrán ser castigados con penas de multa y privación de libertad quienes reprodujeran o plagiaran, en todo o en parte, una obra literaria, artística o científica fijada en cualquier tipo de soporte sin la preceptiva autorización. Ninguna parte de esta publicación, incluido el diseño de la cubierta, puede ser reproducida, almacenada o transmitida de ninguna forma, ni por ningún medio, sea éste electrónico, químico, mecánico, electro-óptico, grabación, fotocopia o cualquier otro, sin la previa autorización escrita por parte de la Editorial.

Otras delegaciones:

México y Centroamérica:
Tel. (52) 281-29-06
Fax (52) 281-29-06
clientes@thomsonlearning.com.mx
clientes@thomsonlearning.com.mx
México, D.F.

Puerto Rico:
Tel. (787) 768-75-80 y 81
Fax (787) 768-75-73
thomson@compul.net
Hato Rey

Chile:
Tel. (562) 521-26-47
Fax (562) 524-46-88
distribucion@thomsonlearning.cl
Santiago

Costa Rica:
EDISA
Tel./Fax (506) 226-88-88
edisa@edisa.com.cr
San José

Colombia:
Tel. (51) 340-64-78
Fax (51) 340-66-75
edisa@edisa.com.co
Bogotá

Cono Sur:
Paseo Santa Rosa, 1041
C.P. 141 - Ciudad de Buenos Aires
Tel. 4833-3038/3833 - 4821-0764
thomson@thomsonlearning.com.ar
Buenos Aires (Argentina)

República Dominicana:
Caribbean Marketing Services
Tel. 809 523-26-27
Fax (809) 523-18-42
cmr@caribbean.net.do

Bolivia:
Librerías Asociadas, S.R.L.
Tel./Fax (591) 2244-53-89
libras@libras.com.bo
La Paz

Paraguay:
Ediciones Remedio
Tel. (592) 780-20-82 y 780-29-71
Fax (592) 780-45-80
ediciones@remedio.net
Caracas

El Salvador:
The Bookshop, S.A. de C.V.
Tel. (503) 242-30-17
Fax (503) 242-13-80
emr@thebookshop.net
San Salvador

Guatemala:
Nactus, S.A.
Tel. (502) 368-01-48
Fax (502) 368-15-70
nactus@nactus.com.gt
Guatemala

Índice General

Prólogo	v
Parte I Funciones Escalares	1
Capítulo 1 Conceptos básicos	3
1.1 Dominio, recorrido y gráfica de una función	- 3
1.2 Geometría del plano y el espacio	7
1.3 Curvas de nivel	42
1.4 Funciones de n variables. El espacio \mathbb{R}^n	52
1.5 Problemas	63
Capítulo 2 Continuidad	65
2.1 Generalidades. Conjuntos abiertos y cerrados	65
2.2 Límite y continuidad de una función	70
2.3 Problemas	81
Capítulo 3 Derivadas parciales. Diferenciabilidad	85
3.1 Introducción	85
3.2 Derivadas parciales	92
3.3 Diferenciabilidad y aproximación lineal	99
3.4 Notas al capítulo	123
3.5 Problemas	131
Capítulo 4 Consecuencias de la diferenciabilidad	135
4.1 La regla de la cadena	136
4.2 Funciones implícitas	162
4.3 Notas al capítulo	196
4.4 Problemas	198

Capítulo 5	Derivadas sucesivas. Fórmula de Taylor	211
5.1	Derivadas sucesivas. Funciones de clase C^p	211
5.2	Funciones compuestas e implícitas	224
5.3	Fórmula de Taylor I	231
5.4	Fórmula de Taylor II	242
5.5	Notas y complementos	254
5.6	Problemas	265
Capítulo 6	Optimización	275
6.1	Máximos y mínimos	275
6.2	Restricciones de igualdad	309
6.3	Restricciones de desigualdad	339
6.4	Problemas	354
Parte II	Funciones Vectoriales	363
Capítulo 7	Topología del espacio \mathbb{R}^n	365
7.1	Conjuntos abiertos y cerrados	366
7.2	Convergencia en \mathbb{R}^n	370
7.3	Conjuntos compactos	374
7.4	Límite y continuidad de una función	377
7.5	Normas en \mathbb{R}^n	390
7.6	Aplicaciones lineales	392
7.7	Conexión	395
7.8	Notas y complementos	401
7.9	Problemas	410
Capítulo 8	Funciones diferenciables	423
8.1	Diferencial de una función	423
8.2	La regla de la cadena y sus consecuencias	428
8.3	Curvas en \mathbb{R}^n	435
8.4	Notas y complementos	445
8.5	Problemas	450

Capítulo 9 El teorema de la función implícita	453
9.1 Aproximaciones sucesivas	453
9.2 Teoremas de la función implícita e inversa	459
9.3 Dependencia funcional	470
9.4 Teoremas de Lagrange y Kuhn-Tucker	473
9.5 Notas y complementos	479
 Parte III El Cálculo Diferencial en la Economía	 481
 Capítulo 10 Nociones de economía marginalista	 483
10.1 La teoría marginalista	483
10.2 Magnitudes marginales y elasticidades	487
10.3 Oferta y demanda: ejemplos	492
 Capítulo 11 Teoría de la oferta y la demanda	 499
11.1 Teoría clásica de la demanda	500
11.2 Teoría clásica de la oferta	518
 Bibliografía	 525
 Índice de Materias	 527

Prólogo

Hay publicados excelentes tratados y multitud de textos de Cálculo Diferencial, por lo que la aparición de un nuevo libro sobre una materia tan establecida requiere, sin duda, algunas explicaciones. Las hay de carácter general: los enfoques cambian (también hay "modas" en matemáticas, no sólo por un prurito de novedad: se aprende de los errores), lo mismo ocurre con las circunstancias (formación de los alumnos que llegan a la Universidad, nuevos desarrollos de las Ciencias aplicadas, nuevos planes de estudios,...) y por ello, como dice el gran místico Miguel de Molinos en la presentación de su *Gufa Espiritual*, "No admire ver salir cada día a la luz del mundo nuevos libros...". Ni todo está dicho ni todo está escrito, y así habrá que escribir hasta el fin del mundo".

Presentamos, pues, un nuevo texto dedicado al Cálculo Diferencial de varias variables concebido y elaborado de forma que pueda ser utilizado tanto en facultades de Ciencias y escuelas técnicas como en carreras en las que esta materia tiene menor carga lectiva, como Biología, Económicas y Empresariales, etc.

Para dar cuenta de las razones concretas que, más allá de las motivaciones generales antes indicadas, han llevado a los autores a escribir este libro, permítansenos algunas consideraciones en torno a la docencia de matemáticas para economistas, campo en el que acumulamos cierta experiencia (como también, desde luego, en la docencia para matemáticos). Indiquemos, de paso, que aunque nos concentremos en la Economía por razones de experiencia y —por qué negarlo— gusto personal, las observaciones que siguen son aplicables a casi todas las carreras que precisan, en mayor o menor medida, de contenidos de tipo matemático.

En los últimos años hemos asistido a un fenómeno curioso: junto a una etapa imparable de matematización de la economía (en muchos casos probablemente excesiva), la carga lectiva de las materias de contenido matemático de las carreras de Económicas y Empresariales se ha reducido de forma considerable. Esto sucede a nivel mundial, pero es particularmente cierto en nuestro país. Este proceso ha ido acompañado, a su vez, de un progresivo abandono

de la metodología característica de las matemáticas, es decir del rigor en los planteamientos iniciales y los sucesivos procesos deductivos que constituyen la "marca de la casa" matemática.

Esta aparente —y no sólo aparente— contradicción tiene varias explicaciones: en el pasado se ha abusado hasta la saciedad de la abstracción matemática en muchos aspectos bastante desconectados del objeto de interés del economista, al cual se le ha hecho pagar el precio de la moda imperante en el momento, y se le ha hecho creer que *todas* las partes de la matemática son igualmente importantes para su disciplina, incluyendo larguísima exposiciones de fundamentos abstractos (teoría de conjuntos, lógica formal,...), sin duda esenciales en el ámbito interno matemático, pero cuyo objeto es exclusivamente el servir de base lógica sólida para el instrumento matemático que *realmente* va a utilizar el economista.

Un ejemplo típico lo tenemos, precisamente, en el Cálculo Diferencial, instrumento de importancia capital para la Teoría Económica. Para entender bien el concepto de derivada es necesario haber establecido previamente el de límite, el cual, a su vez, requiere nociones de topología de los números reales, lo que nos lleva a la construcción formal de éstos mediante aproximaciones racionales que, a su vez, son pares de enteros que se definen en términos del número de elementos de un conjunto. Y ya hemos llegado a la Teoría de Conjuntos. Por tanto, no habrá más remedio que hablar de conjuntos y sus propiedades para ir, paso a paso, desplegando el edificio matemático para llegar a nuestro objetivo inicial: la derivada (con suerte, para finales de enero). Desde el punto de vista matemático, la concatenación lógica antes expuesta es perfecta e irrenunciable (hasta que a alguien se le ocurra otra mejor), pero ha llevado a cometer lo que, en nuestra opinión, es un grave error metodológico: reproducir en las clases de *matemáticas para economistas* —o biólogos, químicos, ingenieros...— lo que serían contenidos típicos de *matemáticas para matemáticos*. Y creemos que el error es grave porque estos contenidos de fundamentación pueden alcanzar niveles de abstracción francamente difíciles de aprehender por el alumno recién llegado a la carrera de Económicas, Empresariales, Biología..., aparte de exigir un tiempo precioso y necesario para exponer al alumno los contenidos y métodos que *realmente* aplicará en su ciencia particular. Ante la imposibilidad (sería además un disparate) de reproducir la mitad de la carrera de Matemáticas en los estudios de Económicas, la respuesta a ese modelo "fundamentalista" antes descrito no se hizo esperar: expliquemos a los alumnos *sólo* lo que realmente van a utilizar, hagamos que aprendan unas cuantas técnicas de análisis gráfico, cálculo diferencial, optimización, etc. Pero esta solución también es, a nuestro parecer, errónea: para estudiar la derivada *sí* se necesita entender algo de límites y la idea de que se puede utilizar una herramienta matemática sin comprender su mecanismo interno (como conducir un coche

sin conocer su mecánica) es válida en determinados contextos muy concretos pero no lo es en general. No tiene sentido —a nivel universitario— exponer la anulación de la derivada (como condición necesaria de óptimo) como si fuese una “regla” más a memorizar, cuando, una vez entendida, esa propiedad es de sentido común y está llena de lógica.

Nuestro libro intenta situarse, en el ámbito de la docencia de matemáticas para economistas y otros científicos aplicados, en un campo intermedio entre los textos que juzgamos demasiado avanzados (por los temas tratados y/o el estilo de exposición) y los que llegan a tales extremos en su objetivo de “simplificación” y “acercamiento al alumno” que más parecen colecciones de fórmulas o manuales de aplicaciones de software. Creemos que, efectivamente, hay muchos contenidos matemáticos muy difíciles de captar por un alumno en sus primeros años de carrera (donde, en las licenciaturas de Económicas y otras afines, se suelen situar los cursos de matemáticas) y consideramos fundamental simplificarlos *sin que pierdan su esencia*. Así, preferimos exponer las ideas básicas para las funciones de dos variables, utilizar gráficas como instrumento metodológico fundamental, recurrir a la intuición cuanto sea posible,... todo ello para lograr que el alumno *entienda* un concepto y lo que hay detrás (o debajo) de él, sin resignarnos a que, simplemente, “lo sepa manejar”.

Precisando algo más, digamos que el libro tiene un triple objetivo representado por las tres partes en que se divide:

- En la primera de ellas se desarrollan los elementos básicos del Cálculo Diferencial de varias variables, con énfasis en la comprensión y procurando una redacción flexible que permita una presentación fluida y sin interrupciones (se reservan, por ejemplo, para el final de cada apartado los conceptos más avanzados y las demostraciones más complicadas) de las ideas fundamentales y las técnicas más importantes (derivación explícita e implícita, desarrollo de Taylor, optimización,...). Todo ello a un nivel *simple pero fundamentado*.
- La segunda parte tiene un carácter más avanzado, tanto en los temas tratados como en el estilo de exposición, y en ella se justifican con rigor los —escasos— resultados difíciles que se han utilizado sin demostración en la primera parte. Son temas que, como decíamos antes, no creemos que sean propios de los primeros años de los estudios de Económicas, pero pueden cubrir ampliamente las necesidades de un curso más formal orientado a la Economía matemática. Por otro lado, con ellos se completan —y *este es otro de los objetivos de este libro*— los contenidos y el nivel con que se imparte la disciplina en las tradicionales carreras científicas y técnicas. En esta parte, el estilo de presentación en cierto modo se “invierte” y se pone el énfasis en la *precisión del razonamiento*.

Se entiende que el lector está ya familiarizado con los conceptos antes de abordar esta segunda parte, aunque en las carreras técnicas se puede utilizar un método de "zigzag" entre las dos partes, incorporando las demostraciones a medida que se van cubriendo los contenidos.

- La tercera parte contiene un breve resumen de las aplicaciones más típicas del cálculo diferencial a la economía. Está dirigida a estudiantes de Económicas y Empresariales, y también a los de otras titulaciones científicas y técnicas cuyos planes de estudios más recientes incluyen asignaturas de Economía. La presentación en bloque independiente tiene la ventaja de permitir su exposición (si procede) en el momento en que el profesor considere oportuno. Por otra parte, sería nuestro deseo que esta sección final sirviese también como estímulo para estudiantes de otras carreras, que pueden hallar en estos capítulos una introducción muy asequible a ciertas aplicaciones de las matemáticas que quizá les sorprendan. Nuestros planes de estudios son muy poco dados a fomentar la curiosidad por contenidos que no estén directamente conectados con la carrera elegida, en matrimonio indisoluble, por el alumno. Y, sin embargo, muchos grandes economistas llegaron a este campo procedentes de otras disciplinas, atraídos por la elegancia de sus planteamientos y desarrollo formal. Al menos, esperamos que los lectores no economistas aprovechen esta parte final como mero objeto de curiosidad y para "hojear de vez en cuando", aunque sólo sea por la satisfacción de saber que "eso no entra en el examen".

Metodología

Cada capítulo se desarrolla con la mayor independencia posible del resto, y se intenta que cada unidad temática (capítulo, apartado,...) avance en nivel de dificultad creciente. Es decir, los contenidos más complejos se encuentran al final de la unidad, de forma que el lector, convenientemente aconsejado, pueda interrumpir su lectura y pasar a la siguiente sin grave ruptura de continuidad. Análogamente, las notas y complementos al final de los capítulos se pueden omitir en primera lectura, y tienen como objetivo estimular el interés sobre aspectos más avanzados de la materia; se trata de ilustrar que hay "todo un mundo" más allá del capítulo que termina, sin "imponerlo" en el texto principal pero ofreciendo un "aperitivo" al alumno motivado.

Los apartados de problemas contienen ejercicios similares a los que se han desarrollado —con detalle— en el cuerpo básico del capítulo, y también problemas más difíciles en los que se guía al alumno mediante indicaciones y se pretende que éste descubra por sí mismo algunas facetas interesantes de la teoría o de la práctica.

Parte I

Funciones Escalares

Capítulo 1

Conceptos básicos

Comenzamos este capítulo con las primeras definiciones y conceptos de la teoría de funciones de varias variables. Enseguida se aprecia que para progresar en ella es necesario un manejo fluido de la geometría del plano y el espacio que se estudia en cursos anteriores. Para comodidad del lector que lo precise, en el apartado 2 se hace un resumen de carácter informal de los resultados geométricos indispensables que, además, sirve para motivar su extensión, en el apartado 4, a espacios de dimensión n cualquiera. En el apartado 3 se aprovechan las posibilidades de representación gráfica de las funciones de dos variables para introducir a un nivel elemental algunas nociones de la teoría de *optimización*, uno de los capítulos importantes de un curso de cálculo diferencial.

1.1 Dominio, recorrido y gráfica de una función

En una primera etapa del cálculo infinitesimal se estudian las *funciones reales de variable real* $f: \mathbb{R} \rightarrow \mathbb{R}$. Se suelen expresar por $y = f(x)$, con x representando la *variable independiente* e y la *variable dependiente*. Una y otra son números reales que, en las aplicaciones, cuantifican magnitudes cuya relación está modelada matemáticamente por la función $y = f(x)$.

También aparecen con frecuencia en las aplicaciones funciones de *varias variables reales*. Por ejemplo, si consideramos el volumen V de un cilindro circular recto, $V = \pi r^2 h$, resulta ser una función de *dos variables*, r , el radio de la base, y h , la altura: para cada par (r, h) de números reales positivos —que definen un determinado cilindro— se tiene un valor de V bien determinado. En Economía, las *funciones de producción* tienen la forma general $Q = Q(K, L)$ cuando se supone que el producto (Q) está determinado por las cantidades empleadas de capital (K) y trabajo (L).

En la primera parte de este libro estudiaremos las funciones $f: \mathbb{R}^n \rightarrow \mathbb{R}$ de n variables reales x_1, \dots, x_n y con valores reales, es decir, las funciones o aplicaciones que a cada n -upla $x = (x_1, \dots, x_n)$ de un cierto subconjunto D de \mathbb{R}^n hacen corresponder un número real $y = f(x_1, \dots, x_n)$. Los dos ejemplos anteriores son funciones de \mathbb{R}^2 en \mathbb{R} . Con frecuencia, y en aras de una mayor claridad, nuestros argumentos se referirán a funciones de *dos* variables, pues podremos ilustrarlos con figuras bi- o tridimensionales. Las denotaremos por $y = f(x_1, x_2)$ o bien $z = f(x, y)$. Los resultados que obtengamos, sin embargo, serán válidos para cualquier número de variables $n \geq 2$.

En la segunda parte del libro estudiaremos las *funciones vectoriales* $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ con $m \geq 1$ (las funciones $f: \mathbb{R}^n \rightarrow \mathbb{R}$ se denominan *funciones escalares* por ser $y = f(x_1, \dots, x_n)$ una magnitud *escalar*, o sea, que se cuantifica mediante una "escala" sobre una recta, como es el caso de la temperatura de un objeto. La velocidad del viento en un punto P del espacio, en cambio, es una magnitud *vectorial* que no sólo tiene "tamaño" sino también "dirección y sentido" (Fig. 1.1) y necesita, para su correcta cuantificación, tres números reales. Por eso, si el espacio de "llegada" de la función f es \mathbb{R}^m con $m > 1$, se dice que es una función vectorial o con valores vectoriales).

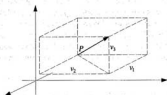


Figura 1.1

Como para las funciones de una variable (y, en general, para las funciones o aplicaciones de un conjunto X en un conjunto Y), el *dominio* (de definición) de una función $f: \mathbb{R}^n \rightarrow \mathbb{R}$ es el conjunto de puntos $x = (x_1, \dots, x_n)$ de \mathbb{R}^n en que, en el contexto de que se trate, está definida la función f ; se denota por $\text{Dom } f$ o $D(f)$. Cuando la función viene dada, como será lo usual en este libro, por una fórmula matemática y no se hace referencia a ninguna aplicación concreta, se suele sobreentender que el dominio consta de todos los puntos (x_1, \dots, x_n) para los que está definida dicha fórmula. Por ejemplo, el dominio de la función

$$z = \sqrt{1 - x^2 - y^2} \quad (1.1)$$

es el círculo del plano $\{(x, y) \in \mathbb{R}^2; x^2 + y^2 \leq 1\}$, centrado en el origen y de

radio 1, pues el radicando ha de ser no negativo para que exista raíz cuadrada (Fig. 1.3). En los dos ejemplos precedentes, la función $V = \pi r^2 h$ está definida para todo $(r, h) \in \mathbb{R}^2$, aunque en el contexto geométrico en que se aplica su dominio "natural" sería el conjunto $\{(r, h) \in \mathbb{R}^2; r > 0, h > 0\}$, o sea, el primer cuadrante del plano sin los ejes, ya que el radio y la altura son, por su propia naturaleza geométrica, magnitudes positivas; y, análogamente, el dominio natural de una función de producción $Q = Q(K, L)$ sería $\{(K, L) \in \mathbb{R}^2; K \geq 0, L \geq 0\}$.

El valor y correspondiente a $x = (x_1, \dots, x_n)$ se dice que es la *imagen* de x por f . El conjunto de todas las imágenes es la *imagen, rango o recorrido* de la función f . Se representa por $\text{Im } f$ o $R(f)$ y es el conjunto de todos los valores que toma la variable dependiente y . Si A es un subconjunto del dominio de f , la *imagen de A por f* es el conjunto

$$f(A) \stackrel{\text{def}}{=} \{y \in \mathbb{R}; y = f(x) \text{ para algún } x \in A\}$$

EJEMPLO. La función $z = f(x, y) = \sqrt{1 - x^2 - y^2}$ verifica claramente $0 \leq f(x, y) \leq 1$. Veamos que *toda* z del intervalo $[0, 1]$ es imagen de puntos (x, y) del dominio de f . Se tiene, para empezar, $f(0, 0) = 1$; dado $\bar{z} \in [0, 1]$, sea $\bar{r} > 0$ tal que $\bar{z} = \sqrt{1 - \bar{r}^2}$; ese número es único y vale $\bar{r} = \sqrt{1 - \bar{z}^2}$; entonces, para todo punto (x, y) de la circunferencia $x^2 + y^2 = \bar{r}^2$ se verifica que $f(x, y) = \bar{z}$. Por tanto, $\text{Im } f = [0, 1]$ (Fig. 1.3). Se intuye por este ejemplo tan simple que, en general, es más difícil determinar el recorrido de una función que su dominio.

El concepto de gráfica de una función de dos variables se obtiene extendiendo de manera natural el correspondiente para funciones de una variable. Se recordará que, si $y = f(x)$ es una función cualquiera de una variable, su gráfica venía dada por

$$\{(x, y) \in \mathbb{R}^2; y = f(x) \text{ para algún } x \in \text{Dom } f\}$$

La gráfica de una función $y = f(x)$ es, en general, una curva en el plano xy (Fig. 1.2); en los cursos elementales de cálculo infinitesimal se aprende a esbozar su aspecto basándose en las propiedades de la función $y = f(x)$ (dominio de definición, posible existencia de asíntotas mediante el cálculo de límites, crecimiento y puntos críticos mediante la derivada primera, concavidad mediante la derivada segunda,...).

La gráfica de una función $z = f(x, y)$ de dos variables es el conjunto de puntos de \mathbb{R}^3

$$\{(x, y, z); z = f(x, y) \text{ para algún } (x, y) \in \text{Dom } f\}$$

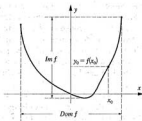


Figura 1.2

y es bastante más complicada de dibujar; en general, se trata de una *superficie* en \mathbb{R}^3 ; una primera ilustración la proporcionan las funciones más sencillas que cabe imaginar, a saber, las *funciones lineales*: como se recordará, las funciones de una variable lineales son las definidas por un polinomio en x de primer grado

$$y = mx + d \quad (1.2)$$

(En rigor, el término de función lineal corresponde a las funciones del tipo $y = mx$, o sea, cuando el “término independiente” d del polinomio es nulo y, en consecuencia, la función verifica la propiedad $f(a_1x_1 + a_2x_2) = a_1f(x_1) + a_2f(x_2)$, cualesquiera que sean $a_1, a_2, x_1, x_2 \in \mathbb{R}$, que caracteriza a las funciones lineales entre espacios vectoriales; a pesar de la ambigüedad que supone, mantendremos el término para designar también a las funciones del tipo (1.2); el contexto dejará claro si nos referimos a ellas o a las lineales en sentido estricto). El dominio de la función (1.2) es toda la recta \mathbb{R} , su recorrido es también todo \mathbb{R} si $m \neq 0$ y su gráfica es una recta del plano.

Del mismo modo, las funciones de dos variables más simples son las funciones lineales

$$z = f(x, y) = ax + by + c \quad (1.3)$$

definidas por un polinomio de primer grado en las variables x e y . El dominio de la función (1.3) es claramente todo el plano \mathbb{R}^2 , y, si a y b no son simultáneamente nulos, el recorrido es todo \mathbb{R} . Su gráfica es un *plano*, la superficie más simple en el espacio tridimensional \mathbb{R}^3 . (Las funciones lineales de n variables son de la forma

$$y = f(x_1, \dots, x_n) = a_1x_1 + \dots + a_nx_n + b \quad (1.4)$$

Su dominio es todo \mathbb{R}^n y, salvo que $f(x_1, \dots, x_n) \equiv b$, el recorrido es todo \mathbb{R} . La gráfica es un subconjunto del espacio \mathbb{R}^{n+1} denominado *hiperplano*.)

Las gráficas de funciones no lineales de dos variables son superficies más difíciles de visualizar; por ejemplo, la gráfica de la función

$$f(x, y) = \sqrt{1 - x^2 - y^2}$$

es la mitad superior de la esfera de centro el origen y radio 1 (Fig. 1.3).

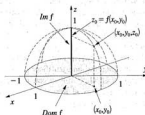


Figura 1.3

Para comprender mejor esta asociación superficie \longleftrightarrow gráfica de función de dos variables, repasaremos al final del apartado siguiente algunas de las superficies del espacio más comunes e importantes. Terminemos este primer apartado con la observación de que para las funciones $y = f(x)$ de una variable y $z = f(x, y)$ de dos variables se tiene que

$$y = f(x) \qquad z = f(x, y)$$

$$\text{Dom } f \subset \mathbb{R} \qquad \text{Dom } f \subset \mathbb{R}^2$$

$$\text{Im } f \subset \mathbb{R} \qquad \text{Im } f \subset \mathbb{R}$$

$$\text{Gra } f \subset \mathbb{R}^2 \qquad \text{Gra } f \subset \mathbb{R}^3$$

1.2 Geometría del plano y el espacio

Se suele decir que la geometría analítica o cartesiana es el estudio de la geometría mediante el álgebra: los puntos se identifican mediante números, las *coordenadas*, y ello permite traducir las propiedades geométricas de las figuras que se desea estudiar a relaciones (*ecuaciones*) entre variables numéricas.

Puntos y vectores Para la asignación de coordenadas hay que elegir un *sistema de referencia*, que ya suponemos *ortogonal*: en el plano, un par de rectas perpendiculares (los *ejes*) con una unidad de medida común (el punto en que se cortan es el *origen de coordenadas*). A cada punto del plano le corresponde el par ordenado de números $(x_1, x_2) \rightarrow (x, y)$ (las *coordenadas rectangulares* o *cartesianas*: *abscisa* y *ordenada*) obtenidos como se indica en la figura 1.4. La correspondencia es “reversible”: elegido un sistema de referencia, a cada par ordenado de números (x_1, x_2) le corresponde un único punto del plano (y así, éste queda identificado al conjunto \mathbb{R}^2 de todos los pares ordenados de números reales y hablaremos de “el punto (x_1, x_2) ”). Otra forma de visualizar el par de números $x = (x_1, x_2)$ es mediante un segmento de recta orientado, a modo de flecha, que parte del origen y termina en el punto $x = (x_1, x_2)$, y cuando se adopta esta representación decimos que $x = (x_1, x_2)$ es un *vector* (Fig. 1.4). Utilizaremos en cada ocasión la interpretación que más nos convenga; con un poco de atención no hay dificultad en identificar cada punto con el vector que parte del origen y termina en él, y ambos con el par ordenado de números (x_1, x_2) . Cuando se desea resaltar el aspecto de vector, con su dirección y sentido, se puede escribir $\vec{x} = (x_1, x_2)$.

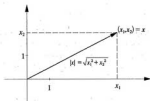


Figura 1.4

Dados dos vectores (puntos) $u = (u_1, u_2)$, $v = (v_1, v_2)$ de \mathbb{R}^2 su suma es el vector

$$u + v = (u_1 + v_1, u_2 + v_2) \quad (1.5)$$

que se obtiene sumando coordenada a coordenada. La suma se realiza gráficamente mediante la *ley del paralelogramo* (Fig. 1.5).

En ocasiones —particularmente en las aplicaciones del *cálculo vectorial*— es útil visualizar un vector $u = (u_1, u_2)$ comenzando en un punto cualquiera (x_1, x_2) y terminando en el punto $(x_1 + u_1, x_2 + u_2)$ (Figura 1.6).

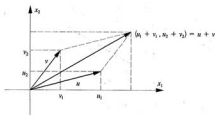


Figura 1.5

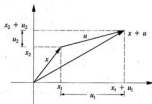


Figura 1.6

Si $c \in \mathbb{R}$ y $u = (u_1, u_2)$, se define

$$cu = (cu_1, cu_2) \quad (1.6)$$

operación que se denomina *multiplicación* (de vectores) *por escalares* (números de \mathbb{R}) (Fig. 1.7).

Se escribe $v - u$ para indicar el vector $v + (-1)u$ (Fig. 1.8).

Dados dos vectores $u = (u_1, u_2)$, $v = (v_1, v_2)$ y dos números reales s y t , el vector $su + tv$ se dice que es una *combinación lineal* de los vectores u y v (Fig. 1.9). Dados dos vectores u y v que no sean uno múltiplo escalar del otro (en cuyo caso se dice que son *linealmente independientes*) cualquier otro vector del plano puede expresarse como combinación lineal de u y v (Fig. 1.9); se dice por ello que constituyen una *base* del conjunto de vectores del plano; la *base canónica* es la formada por los vectores $(1, 0)$ y $(0, 1)$, que, en un sistema de referencia ortogonal, se suponen "colocados" en los ejes perpendiculares de abscisas y ordenadas.

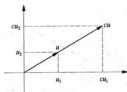


Figura 1.7

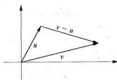


Figura 1.8

Rectas Una recta en el plano está determinada por dos puntos $p = (p_1, p_2)$ y $q = (q_1, q_2)$ (Fig. 1.10).

Sus puntos (x, y) vienen dados por la expresión

$$(x, y) = p + t(q - p), \quad t \in \mathbb{R} \quad (1.7)$$

o bien

$$(x, y) = (1 - t)p + tq, \quad t \in \mathbb{R} \quad (1.8)$$

Dos rectas $r \equiv \{p + t(q - p)\}$ y $s \equiv \{\bar{p} + t(\bar{q} - \bar{p})\}$ son paralelas si los vectores $q - p$ y $\bar{q} - \bar{p}$ son uno múltiplo escalar del otro.

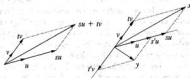


Figura 1.9

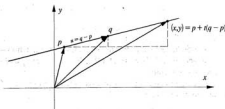


Figura 1.10

Poniendo $u = q - p$, (1.7) toma la forma

$$(x, y) = p + tu, \quad t \in \mathbb{R} \quad (1.9)$$

que pone de manifiesto como una recta queda también determinada por un punto p por el que pasa y un vector dirección $u \neq (0, 0)$ dados. (1.9) (o (1.7)) es la ecuación vectorial de la recta. Desplegándola en coordenadas tenemos las ecuaciones paramétricas de la recta:

$$\begin{cases} x = p_1 + tu_1 \\ y = p_2 + tu_2 \end{cases} \quad (1.10)$$

Eliminando el parámetro t en (1.10) se obtiene la ecuación de la recta en la forma

$$\frac{x - p_1}{u_1} = \frac{y - p_2}{u_2} \quad (1.11)$$

que traduce su caracterización como recta que pasa por el punto (p_1, p_2) y es paralela al vector (u_1, u_2) (Fig. 1.10; si alguna de las coordenadas u_1, u_2 es igual a cero se entiende que el correspondiente numerador en (1.11) también vale cero).

Reordenando los términos de (1.11) se llega a la ecuación implícita de la recta:

$$ax + by + c = 0 \quad (1.12)$$

en donde a y b no son simultáneamente nulos ($(u_1, u_2) \neq (0, 0)$, pues p y q son puntos distintos). Cuando $b \neq 0$ en (1.12) (o sea, cuando la recta no es paralela al eje y) se puede despejar la variable y y la ecuación se puede expresar en forma explícita:

$$y = mx + h \quad (1.13)$$

m es la pendiente de la recta y h la ordenada en el origen. Obsérvese cómo, cuando $b \neq 0$, la ecuación de la recta (1.12) define a y como función (lineal) de x , lo que no ocurre cuando $b = 0$.

El segmento (de recta) que une los puntos $p = (p_1, p_2)$ y $q = (q_1, q_2)$ es el conjunto de puntos del plano

$$[p, q] = \{(1-t)p + tq, t \in [0, 1]\} \quad (1.14)$$

o, equivalentemente,

$$[p, q] = \{p + t(q - p), t \in [0, 1]\} \quad (1.15)$$

(Fig.1.11)

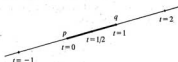


Figura 1.11

Un conjunto $S \subset \mathbb{R}^2$ se dice que es *convexo* si el segmento de recta que une dos cualesquiera de sus puntos está contenido en S (Fig.1.12).

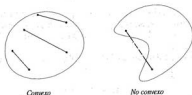


Figura 1.12

NOTA. Hemos considerado desde el principio un sistema de referencia ortogonal, pero para los conceptos que hemos repasado hasta aquí —y los resultados que de ellos se derivarían

sobre pertenencia de un punto a una recta y paralelismo e intersección de rectas, que son relaciones afines— podríamos tomar una referencia cualquiera (Fig. 1.9), pues dichas relaciones se conservan con cualquier cambio de referencia. Para lo que sigue, basado en la noción de distancia en el plano, ya se hace necesaria una referencia ortogonal.

Distancia, producto escalar y ortogonalidad La *distancia* entre dos puntos $p = (p_1, p_2)$ y $q = (q_1, q_2)$ es, de acuerdo con el teorema de Pitágoras (Fig. 1.13),

$$d(p, q) = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2}$$

Nótese que también

$$d(p, q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2} = d(q, p)$$

El *módulo*, *longitud* o *norma* de un vector $u = (u_1, u_2)$ es la cantidad

$$|u| = \sqrt{u_1^2 + u_2^2}$$

o sea, la distancia del punto (u_1, u_2) al origen (Fig. 1.4).

Un vector de norma igual a 1 se dice que es un vector unitario. Obsérvese que (Fig. 1.13)

$$d(p, q) = |p - q| = |q - p|$$

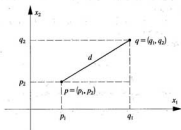


Figura 1.13

Siendo θ el ángulo que forman las rectas determinadas por los vectores p y q (véase la figura 1.14), la fórmula del coseno de la trigonometría establece que

$$|p - q|^2 = |p|^2 + |q|^2 - 2|p||q|\cos\theta \quad (1.16)$$

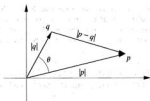


Figura 1.14

Desarrollando $|p - q|^2$ se tiene

$$|p - q|^2 = (p_1 - q_1)^2 + (p_2 - q_2)^2 = |p|^2 + |q|^2 - 2(p_1 q_1 + p_2 q_2)$$

de donde

$$p_1 q_1 + p_2 q_2 = |p| |q| \cos \theta \quad (1.17)$$

La cantidad

$$(p, q) \equiv p \cdot q = p_1 q_1 + p_2 q_2 \quad (1.18)$$

se denomina *producto escalar* (o *interior*) de los vectores p y q . Vemos que dos vectores son perpendiculares (u *ortogonales*) — $\cos \theta = 0$ — si y sólo si su producto escalar es nulo.

Las rectas $r \equiv \{p + tu\}$ y $s \equiv \{q + tv\}$ son perpendiculares si los vectores u y v son ortogonales.

Obsérvese que

$$(u, u) = |u|^2 \quad (1.19)$$

para todo vector $u = (u_1, u_2) \in \mathbb{R}^2$. Se comprueban inmediatamente las siguientes propiedades esenciales del producto escalar (1.18)

$$(u, v) = (v, u) \quad (\text{es simétrico}) \quad (1.20)$$

$$(u + v, w) = (u, w) + (v, w) \quad (\text{es bilineal}) \quad (1.21)$$

$$(u, u) \geq 0 \text{ y } (u, u) = 0 \text{ si y sólo si } u = 0 \quad (\text{es definido positivo}) \quad (1.22)$$

Puesto que $|\cos \theta| \leq 1$, se tiene también

$$|(u, v)| \leq |u| |v| \quad (1.23)$$

que es la *desigualdad de Cauchy-Schwarz*, teniéndose la igualdad si y sólo si u es múltiplo escalar de v o si alguno de los vectores es el vector cero $(0, 0)$. En efecto, si u no es múltiplo escalar de v se tiene $|\cos \theta| < 1$, y si u es múltiplo escalar de v , entonces $\theta = 0$ o $\theta = \pi$ y $|\cos \theta| = 1$.

A partir de (1.20), (1.21), (1.22) y (1.23) se obtienen las propiedades esenciales de la norma de vectores:

$$|u| \geq 0 \text{ para todo } u \in \mathbb{R}^2 \text{ y } |u| = 0 \text{ si y sólo si } u = 0 \quad (1.24)$$

$$|u + v| \leq |u| + |v| \text{ (desigualdad triangular)} \quad (1.25)$$

$$|cu| = |c| |u|, \quad c \in \mathbb{R} \quad (1.26)$$

(Para la desigualdad triangular se parte de

$$|u + v|^2 = \langle u + v, u + v \rangle = |u|^2 + |v|^2 + 2\langle u, v \rangle$$

y se utiliza la desigualdad de Cauchy-Schwarz.) La *desigualdad triangular* para la distancia en el plano es, a partir de (1.25),

$$d(p, z) \leq d(p, q) + d(q, z) \quad (1.27)$$

(Ilústrese (1.25) mediante una figura que justifique el término "desigualdad triangular").

Dada la recta de ecuación

$$ax + by + c = 0 \quad (1.28)$$

sea (x_0, y_0) uno de sus puntos. Entonces

$$ax_0 + by_0 + c = 0 \quad (1.29)$$

Restando (1.29) de (1.28) se obtiene

$$a(x - x_0) + b(y - y_0) = 0 \quad (1.30)$$

es decir, que el vector (a, b) es perpendicular a dicha recta (Fig. 1.15).

Gráficas de funciones no lineales Como decíamos al principio, el modelo del plano cartesiano permite tratar algebraicamente nociones geométricas. El caso de las rectas es el más sencillo: se pueden obtener sus propiedades (pertenencia de puntos a rectas, posiciones relativas entre ellas, etcétera) manejando

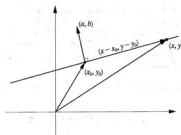


Figura 1.15

las ecuaciones que hemos repasado en lo que antecede. Análogamente, un punto (x, y) del plano estará en la *circunferencia* de centro (x_0, y_0) y radio r si su distancia a (x_0, y_0) vale r , o sea, si

$$\sqrt{(x - x_0)^2 + (y - y_0)^2} = r$$

o, equivalentemente,

$$(x - x_0)^2 + (y - y_0)^2 = r^2 \quad (1.31)$$

que es la *ecuación* de dicha *circunferencia*.

Otras curvas interesantes del plano son las *cónicas*: *elipse*, *parábola* e *hipérbola*, curvas que se obtienen al cortar un cono (doble) por un plano según se muestra en la figura 1.16.

Una segunda vía, equivalente a la anterior, para definir las cónicas es como lugares geométricos de puntos del plano que satisfacen determinadas propiedades geométricas. Así, como se recordará, una *elipse* es el conjunto de puntos del plano tales que la suma de distancias a dos puntos fijos (*focos*) es constante (Fig. 1.17).

La recta que une a los focos corta a la elipse en los *vértices* de ésta. La cuerda que une los vértices es el *eje mayor* de la elipse y su punto medio el *centro*. El *eje menor* es la cuerda perpendicular al eje mayor que pasa por el centro (Fig. 1.17).

La *ecuación canónica* de la elipse se obtiene cuando traducimos la propiedad geométrica que la define a la relación algebraica que satisfacen las coordenadas (x, y) de sus puntos respecto al sistema de referencia que se obtiene situando el eje x según el eje mayor de la elipse y el eje y según el eje menor, con el origen de coordenadas en el centro de la elipse: si los focos están en $(c, 0)$ y

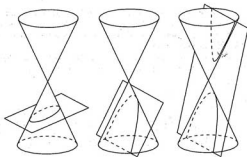


Figura 1.16

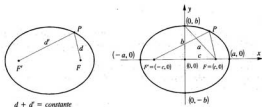


Figura 1.17

$(-c, 0)$ y la suma de las distancias de $P = (x, y)$ a $F = (c, 0)$ y $F' = (-c, 0)$ ha de valer la constante $2a$ (nótese que $c < a$, pues se sobreentiende que $2a$ es mayor que la distancia entre los focos, que vale $2c$), entonces

$$\sqrt{(x-c)^2 + (y-0)^2} + \sqrt{(x+c)^2 + (y-0)^2} = 2a \quad (1.32)$$

Elevando al cuadrado y operando apropiadamente se obtiene

$$\frac{x^2}{a^2} + \frac{y^2}{a^2 - c^2} = 1$$

y con $b^2 = a^2 - c^2$, $b > 0$,

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1 \quad (1.33)$$

que es la ecuación canónica de la elipse. (a es la longitud del semieje mayor y b la del semieje menor. Si en (1.33) $b > a$, la elipse tiene los focos sobre el eje y . Si $a = b$, los focos se confunden en el centro y se tiene como caso particular una circunferencia).

Una *hipérbola* es el conjunto de puntos del plano tales que la diferencia de sus distancias a dos puntos fijos (*focos*) es constante (Fig. 1.18).

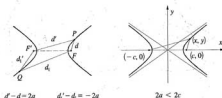


Figura 1.18

La recta que pasa por los focos F y F' corta a las dos ramas de la hipérbola en los vértices. El segmento de recta que une a los vértices es el *eje transversal* y su punto medio el *centro* de la hipérbola.

Si situamos el origen de coordenadas en el centro y el eje x según el eje transversal, la condición geométrica de la hipérbola se traduce en la relación algebraica (Fig. 1.18)

$$\sqrt{(x-c)^2 + (y-0)^2} - \sqrt{(x+c)^2 + (y-0)^2} = \pm 2a$$

Operando se obtiene la ecuación canónica de la hipérbola:

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1 \quad (1.34)$$

donde $b^2 = c^2 - a^2$, $b > 0$. La hipérbola de ecuación (1.34) tiene dos *asíntotas*, las rectas $y = \frac{b}{a}x$ e $y = -\frac{b}{a}x$. El segmento que une los puntos $(0, b)$ y $(0, -b)$ es el *eje conjugado* de la hipérbola. Si los focos están sobre el eje y , la ecuación canónica es

$$\frac{y^2}{a^2} - \frac{x^2}{b^2} = 1 \quad (1.35)$$

Una *parábola* es el lugar geométrico de los puntos del plano que equidistan de una recta fija (*directriz*) y un punto fijo (*foco*) que no está en dicha recta. El punto medio entre el foco y la directriz es el *vértice* de la parábola y la recta que pasa por el foco y el vértice, el *eje* (Fig. 1.19)

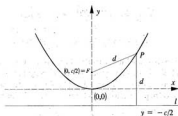


Figura 1.19

Si introducimos unos ejes cartesianos con el origen en el vértice y el eje y según el eje de la parábola de forma que el foco sea $F = (0, c/2)$ y la directriz la recta de ecuación $y = -c/2$, la condición geométrica se traduce en

$$\sqrt{x^2 + \left(y - \frac{c}{2}\right)^2} = y + \frac{c}{2}$$

de donde se obtiene la ecuación canónica

$$x^2 - 2cy = 0 \quad (1.36)$$

(Si $c < 0$, con el foco por debajo de la directriz, la parábola se abre hacia abajo).

Si se cambian los papeles entre los ejes x e y , eligiendo como eje de la parábola el eje x , se tiene la ecuación canónica

$$2cx - y^2 = 0 \quad (1.37)$$

La ecuación (que satisfacen las coordenadas de los puntos) de cualquier lugar geométrico depende naturalmente de la posición de los ejes coordenados. Para obtener la ecuación canónica (o sea, la ecuación más simple) de cada una de las cónicas hemos hecho una elección especial de sistema de referencia. Pero, fijado un sistema de referencia ¿cómo será la ecuación de una elipse —o una parábola, o una hipérbola— cualquiera, no colocada en su posición canónica? La respuesta es que la ecuación de cualquier cónica es una ecuación de la forma

$$ax^2 + 2bxy + cy^2 + 2dx + 2ey + f = 0 \quad (1.38)$$

en la que el primer miembro es un polinomio de segundo grado en las variables x e y ; ejemplos simples son:

$$\begin{array}{ll} \frac{(x-x_0)^2}{a^2} + \frac{(y-y_0)^2}{b^2} = 1 & \text{elipse} \\ xy = k & \text{hipérbola} \\ y = ax^2 + bx + c & \text{parábola} \end{array}$$

Y lo notable es que también es cierto el recíproco, a saber, que la gráfica de una ecuación (1.38) de segundo grado (o sea, el conjunto de puntos del plano cuyas coordenadas satisfacen la ecuación (1.38)) es, salvo casos degenerados, una cónica; en efecto, partiendo de las formas matriciales de la ecuación (1.38):

$$(x, y, 1) \begin{pmatrix} a & b & d \\ b & c & e \\ d & e & f \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = 0 \quad (1.39)$$

y

$$(x, y) \begin{pmatrix} a & b \\ b & c \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + 2(x, y) \begin{pmatrix} d \\ e \end{pmatrix} + f = 0 \quad (1.40)$$

se puede hallar un sistema de referencia especial (realizando en el sistema de referencia dado un giro seguido de una traslación) respecto del cual el lugar geométrico en cuestión tiene una ecuación sencilla, la *ecuación canónica* ((1.33), (1.34) y (1.36) en los casos más importantes), que permite *clasificarlo* (véase [8] para un estudio detallado de las cónicas en el plano). Los distintos tipos de gráficas de la ecuación (1.38) están determinados por los determinantes

$$D = \begin{vmatrix} a & b & d \\ b & c & e \\ d & e & f \end{vmatrix} \quad \text{y} \quad D_0 = \begin{vmatrix} a & b \\ b & c \end{vmatrix}$$

Si $D = 0$ se trata de una *cónica degenerada*. Según el signo de D_0 se tiene:

- $D = 0, D_0 < 0$: dos rectas que se cortan (Ej.: $(x+y-2)(x+2y)=0$)
- $D = 0, D_0 > 0$: un punto (Ej.: $x^2+y^2=0$)
- $D \neq 0, D_0 = 0$: el conjunto vacío (Ej.: $x^2+1=0$), una recta (Ej.: $(x+y-2)^2=0$) o dos rectas paralelas (Ej.: $(x+y-2)(x+y)=0$).

Si $D \neq 0$ se trata de una *cónica no degenerada*, teniéndose:

- $D \neq 0, D_0 > 0$: una elipse (o el conjunto vacío, por ejemplo $x^2+y^2+1=0$)

- $D \neq 0, D_0 < 0$: una hipérbola
- $D \neq 0, D_0 = 0$: una parábola

Hemos ido repasando las ecuaciones de rectas, circunferencias y cónicas. En general, la ecuación de un lugar geométrico del plano será una expresión de la forma

$$F(x, y) = 0 \quad (1.41)$$

que han de satisfacer las coordenadas (x, y) de todos sus puntos. Visto desde el otro ángulo, dada una función $F(x, y)$ de dos variables, la gráfica de la ecuación $F(x, y) = 0$ es el lugar geométrico de los puntos del plano cuyas coordenadas (x, y) satisfacen $F(x, y) = 0$. El estudio de rectas y cónicas se puede realizar de una manera muy completa mediante argumentos puramente algebraicos y geométricos. Pero para estudiar la gráfica de una ecuación general $F(x, y) = 0$ se necesita recurrir al cálculo infinitesimal. Una primera cuestión es saber si la ecuación $F(x, y) = 0$ realmente define implícitamente a y como función de x , o sea, si existe una función $y = f(x)$ tal que

$$F(x, f(x)) = 0 \quad (1.42)$$

para x moviéndose en un cierto intervalo I . Por ejemplo, la ecuación de la circunferencia

$$x^2 + y^2 = 1$$

define dos funciones

$$\begin{aligned} y &= \sqrt{1 - x^2} \\ y &= -\sqrt{1 - x^2} \end{aligned}$$

ambas definidas para $x \in [-1, 1]$, cuyas gráficas son, respectivamente, la semicircunferencia superior y la semicircunferencia inferior. En cambio, la ecuación

$$x^2 + y^2 + 1 = 0$$

no define ninguna función $y = f(x)$ (ni tampoco $x = g(y)$ ¿por qué?).

Si la respuesta a aquella cuestión es positiva, los puntos $(x, f(x))$, $x \in I$, están en la gráfica de $F(x, y) = 0$ y constituyen, por definición, la gráfica de la función $y = f(x)$. Ésta es, bajo hipótesis muy generales sobre F , una curva del plano que aprendemos a esbozar en los cursos elementales de cálculo

infinitesimal estudiando las propiedades de la función $y = f(x)$ (dominio de definición, posible existencia de asíntotas mediante el cálculo de límites, crecimiento y puntos críticos mediante la derivada primera, concavidad mediante la derivada segunda,...).

La ecuación $F(x, y) = 0$ puede venir dada en términos de operaciones simples como en el ejemplo anterior, y entonces permite “despejar” fácilmente la variable y en función de la variable x . O bien se trata ya de la forma explícita $y = f(x)$ de una función, como

$$y = 3x^5 + 2x$$

que se puede interpretar como definida por la ecuación implícita

$$3x^5 + 2x - y = 0$$

siendo trivial en este caso el paso de la ecuación (forma implícita) a la forma explícita de la función. Pero en otros casos, como la ecuación

$$x + \log x + y + \log y - 2 = 0$$

no está claro si estará o no definida implícitamente una función $y = f(x)$. En los capítulos 4 y 9 se analizará esta cuestión, que constituye uno de los temas más importantes de un curso de Cálculo Diferencial.

Coordenadas polares Los puntos del plano se pueden describir mediante otros sistemas de coordenadas (es decir, mediante pares de números) diferentes del de coordenadas rectangulares. Después de éste, el más útil es el sistema de coordenadas polares, en el cual se asigna a cada punto del plano un par de números de la manera siguiente: elegidos un punto O del plano (el polo) y una semirrecta con una unidad de medida que tiene por origen dicho punto (el eje polar), cada punto P del plano —distinto del polo O — queda determinado por dos números: la distancia r entre O y P y el ángulo $\theta \in [0, 2\pi)$ (en radianes) formado por el eje polar y la semirrecta con origen en O que pasa por P , considerando como positivo el sentido contrario a las agujas del reloj (figura 1.20). El par de números (r, θ) son las coordenadas polares de P ; r es el radio vector y θ el argumento de P (al punto O no se le asigna valor de θ ; se entiende, simplemente, que se corresponde con el valor $r = 0$; obsérvese que, desde el punto de vista geométrico, en el sistema de coordenadas rectangulares un punto queda descrito como la intersección de una recta vertical con una recta horizontal, mientras que en el de coordenadas polares, como la intersección de una circunferencia con una semirrecta que parte del centro de aquella). Si (r, θ) son coordenadas polares de P , también lo son $(r, \theta + 2k\pi)$ con k entero. Si se desea un único par de números (r, θ) para cada punto

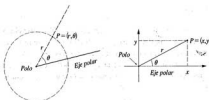


Figura 1.20

$P \neq O$, se entenderá que $\theta \in [0, 2\pi)$ (en ocasiones, puede convenir el intervalo $(-\pi, \pi]$ para el recorrido de θ).

Se suele hacer coincidir el polo con el origen de las coordenadas cartesianas y el eje polar con el semieje positivo de abscisas (Fig. 1.20). En ese caso, dadas las coordenadas polares (r, θ) de un punto, las coordenadas cartesianas vienen dadas por

$$x = r \cos \theta, \quad y = r \sin \theta$$

A su vez, si (x, y) son las coordenadas cartesianas, entonces

$$r = \sqrt{x^2 + y^2}$$

y $\theta \in [0, 2\pi)$ es tal que

$$\cos \theta = \frac{x}{\sqrt{x^2 + y^2}}, \quad \sin \theta = \frac{y}{\sqrt{x^2 + y^2}}$$

(para $x \neq 0$ se tiene

$$\theta = \begin{cases} \arctan(y/x) & \text{si } x > 0, y \geq 0 \\ \pi + \arctan(y/x) & \text{si } x < 0 \\ 2\pi + \arctan(y/x) & \text{si } x > 0, y < 0 \end{cases}$$

donde se entiende que se toma la determinación principal de la función arco tangente, es decir, $\arctan(y/x) \in (-\pi/2, \pi/2)$. Si $x = 0$, entonces $\theta = \pi/2$ si $y > 0$ y $\theta = 3\pi/2$ si $y < 0$).

Como antes, se pueden describir lugares geométricos del plano mediante ecuaciones satisfechas por sus coordenadas polares. Por ejemplo, la ecuación de una circunferencia de centro el origen (el polo) y radio r_0 es simplemente $r = r_0$. (Ejercicio: a) comprobar que la ecuación $r = 2 \sin \theta$ representa una circunferencia; determinar su centro y su radio. b) representar gráficamente la ecuación $r = 2\theta$, entendiendo aquí que $\theta \in [0, \infty)$.)

El espacio \mathbb{R}^3

Puntos y vectores En el espacio, un sistema de referencia ortogonal está formado por tres rectas (los ejes) que se cortan en un punto (el origen) y son perpendiculares dos a dos, con una unidad de medida común. A cada punto del espacio le corresponde una terna ordenada de números $(x_1, x_2, x_3) \rightarrow (x, y, z)$ —(las *coordenadas cartesianas o rectangulares*) como se indica en la figura 1.21.

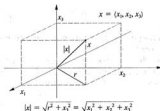


Figura 1.21

Como en el plano, la correspondencia es reversible, de manera que el conjunto \mathbb{R}^3 de ternas ordenadas de números reales (x_1, x_2, x_3) queda identificado al espacio tridimensional. También es la misma la noción de vector y de suma de vectores y multiplicación por escalares:

$$u + v = (u_1, u_2, u_3) + (v_1, v_2, v_3) = (u_1 + v_1, u_2 + v_2, u_3 + v_3) \quad (1.43)$$

$$cu = (cu_1, cu_2, cu_3) \quad (1.44)$$

así como su interpretación gráfica (si bien ésta es, lógicamente, algo más complicada en tres que en dos dimensiones).

Rectas La ecuación vectorial de la recta determinada por los puntos $p = (p_1, p_2, p_3)$ y $q = (q_1, q_2, q_3)$ es

$$(x, y, z) = p + t(q - p), \quad t \in \mathbb{R} \quad (1.45)$$

o equivalentemente

$$(x, y, z) = p + tu, \quad t \in \mathbb{R} \quad (1.46)$$

donde $u = q - p$. (Fig. 1.22)

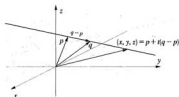


Figura 1.22

Al desplegar (1.46) en coordenadas se tienen las *ecuaciones paramétricas* de la recta

$$\begin{cases} x = p_1 + tu_1 \\ y = p_2 + tu_2 \\ z = p_3 + tu_3 \end{cases} \quad (1.47)$$

Eliminando t en (1.47) se tiene la *forma continua* de la ecuación de la recta:

$$\frac{x - p_1}{u_1} = \frac{y - p_2}{u_2} = \frac{z - p_3}{u_3} \quad (1.48)$$

(si alguna de las coordenadas u_1, u_2, u_3 es nula, se entiende que también lo es el correspondiente numerador).

Las definiciones de rectas paralelas, segmento y conjunto convexo son las mismas que en el plano. (En el plano, dos rectas distintas o son paralelas o se cortan; en el espacio hay una tercera posibilidad: que se crucen, lo que ocurre cuando no están contenidas en un mismo plano.)

Planos Los planos en el espacio vienen definidos en términos de los conceptos previos de punto y recta. Dados un punto p y una recta r que no contiene a p , el plano que determinan es el conjunto de rectas que pasan por p y por los distintos puntos de r , más la recta paralela a r que pasa por p (Fig. 1.23).

Si $r = \{q + tv; t \in \mathbb{R}\}$, los puntos del plano estarán dados, con λ recorriendo \mathbb{R} , por $(x, y, z) = p + \lambda(q + tv - p)$, o sea, poniendo $u = q - p$ y $\mu = \lambda t$, por

$$(x, y, z) = p + \lambda u + \mu v \quad (1.49)$$

Con $\lambda = 0$ y μ arbitrario se incluye en (1.49) la recta paralela a r que pasa por p . (1.49) muestra al plano determinado por un punto del mismo y

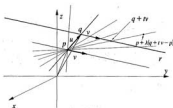


Figura 1.23

dos vectores que no son uno múltiplo escalar del otro (Fig. 1.24). (Equivalentemente, un plano queda determinado por tres puntos del espacio no alineados en una recta: $p, p + u, p + v$).

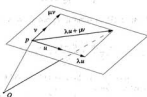


Figura 1.24

La ecuación (1.49), con λ y μ recorriendo \mathbb{R} , es la ecuación vectorial del plano. Cuando p es el origen de coordenadas, el plano pasa por el origen y es el conjunto de vectores que son combinación lineal de los vectores u y v :

$$\{\lambda u + \mu v; \lambda, \mu \in \mathbb{R}\}$$

o sea, lo que se conoce como *subespacio vectorial engendrado por los vectores u y v* . Dos planos $\{p + \lambda u + \mu v\}$ y $\{\tilde{p} + \lambda \tilde{u} + \mu \tilde{v}\}$, son *paralelos* si los subespacios vectoriales $\{\lambda u + \mu v\}$ y $\{\lambda \tilde{u} + \mu \tilde{v}\}$ coinciden. Todo plano de \mathbb{R}^3 es el “trasladado paralelamente” de un plano que pasa por el origen, su subespacio vectorial —de dimensión dos— asociado (denominado *variedad de dirección del plano*).

Desplegando (1.49) coordenada a coordenada se obtienen las ecuaciones

paramétricas del plano:

$$\begin{cases} x = p_1 + \lambda u_1 + \mu v_1 \\ y = p_2 + \lambda u_2 + \mu v_2 \\ z = p_3 + \lambda u_3 + \mu v_3 \end{cases} \quad (1.50)$$

Eliminando λ y μ en (1.50) se obtiene la ecuación implícita del plano:

$$ax + by + cz + d = 0 \quad (1.51)$$

La eliminación puede llevarse a cabo despejando y sustituyendo (por ejemplo, primero se despeja λ en la primera ecuación de (1.50) y se sustituye en la segunda y la tercera; quedarán dos ecuaciones en μ y se utiliza una de ellas para despejar μ ; se sustituye en la otra y queda finalmente una ecuación en x, y, z sin parámetros). Pero resulta más transparente argumentando como sigue: el punto (x, y, z) pertenece al plano si y sólo si el sistema lineal en λ y μ (1.50) es compatible y determinado, o sea, si y sólo si

$$\begin{vmatrix} x - p_1 & u_1 & v_1 \\ y - p_2 & u_2 & v_2 \\ z - p_3 & u_3 & v_3 \end{vmatrix} = 0 \quad (1.52)$$

Desarrollando este determinante por la primera columna obtenemos

$$\begin{vmatrix} u_2 & v_2 \\ u_3 & v_3 \end{vmatrix} (x - p_1) - \begin{vmatrix} u_1 & v_1 \\ u_3 & v_3 \end{vmatrix} (y - p_2) + \begin{vmatrix} u_1 & v_1 \\ u_2 & v_2 \end{vmatrix} (z - p_3) = 0$$

y, operando y simplificando,

$$ax + by + cz + d = 0$$

con a, b, c no todos nulos pues los vectores u y v no son uno múltiplo escalar del otro (o sea, son linealmente independientes).

En general, lo mismo que decíamos para el plano \mathbb{R}^2 , la ecuación de un lugar geométrico del espacio será una expresión de la forma

$$F(x, y, z) = 0$$

que han de satisfacer las coordenadas (x, y, z) de todos sus puntos. A su vez, dada una función $F(x, y, z)$ de tres variables, la gráfica de la ecuación $F(x, y, z) = 0$ es el lugar geométrico de los puntos cuyas coordenadas (x, y, z) satisfacen la relación $F(x, y, z) = 0$.

Si la función F es lineal: $F(x, y, z) = ax + by + cz + d$, y no es $F \equiv d$, la correspondiente ecuación es de la forma (1.51) y su gráfica es un plano; en efecto, a partir de la ecuación (1.51) podemos escribir unas ecuaciones

paramétricas para la gráfica tomando, por ejemplo, $x = \lambda$, $y = \mu$, si $c \neq 0$, (con las modificaciones apropiadas si $c = 0$) y, a partir de ellas, la ecuación vectorial que responde a la definición original de plano que hemos dado aquí. Tenemos así una caracterización analítica equivalente de la noción de plano: un plano en el espacio es la gráfica de una ecuación lineal en tres variables.

Cuando $c \neq 0$ en la ecuación (1.51), se puede despejar la variable z y se tiene la ecuación explícita del plano

$$z = Ax + By + C \quad (1.53)$$

z es función lineal de las variables x y y . Vemos, como ya señalamos al final del primer apartado, que la gráfica de una función lineal de dos variables es un plano en el espacio.

Distancia, producto escalar y ortogonalidad Aplicando dos veces el teorema de Pitágoras (Fig. 1.25) se tiene que la distancia entre dos puntos $p = (p_1, p_2, p_3)$ y $q = (q_1, q_2, q_3)$ es

$$d(p, q) = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2 + (q_3 - p_3)^2} \quad (1.54)$$

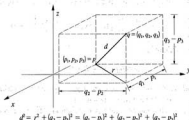


Figura 1.25

Como en \mathbb{R}^2 , la longitud o norma de un vector $u = (u_1, u_2, u_3)$ es la cantidad

$$|u| = \sqrt{u_1^2 + u_2^2 + u_3^2} \quad (1.55)$$

o sea, la distancia del punto (u_1, u_2, u_3) al origen. Se tiene

$$d(p, q) = |p - q| = |q - p|$$

y asimismo

$$p_1q_1 + p_2q_2 + p_3q_3 = |p| |q| \cos \theta \quad (1.56)$$

siendo θ el ángulo que forman los vectores p y q en el plano que determinan pasando por el origen.

El producto escalar (o interior) de p y q es la cantidad

$$\langle p, q \rangle \equiv p \cdot q = p_1q_1 + p_2q_2 + p_3q_3 = |p| |q| \cos \theta \quad (1.57)$$

y verifica, naturalmente, las propiedades (1.20), (1.21), (1.22) y (1.23). Dos vectores p y q de \mathbb{R}^3 son ortogonales si y sólo si $\langle p, q \rangle = 0$.

A su vez, la norma (1.55) verifica las propiedades (1.24), (1.25) y (1.26).

Dado el plano de ecuación

$$ax + by + cz + d = 0 \quad (1.58)$$

sea $p = (p_1, p_2, p_3)$ uno de sus puntos. Entonces

$$ap_1 + bp_2 + cp_3 + d = 0 \quad (1.59)$$

Restando (1.59) de (1.58) se obtiene

$$a(x - p_1) + b(y - p_2) + c(z - p_3) = 0 \quad (1.60)$$

es decir, que el vector (a, b, c) es perpendicular a todas las rectas que pasan por p y están contenidas en el plano (Fig. 1.26) ($\vec{n} = (a, b, c)$ —o cualquier otro de sus múltiplos escalares que sea distinto del vector cero— es un vector normal o característico del plano; una recta r es perpendicular a un plano π si su vector dirección es un vector normal al plano π). Hemos llegado así a

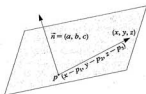


Figura 1.26

otra de las caracterizaciones en geometría euclídea de un plano que podríamos haber tomado como punto de partida: un plano es el conjunto de todas las rectas que pasan por un punto dado p y son ortogonales a una recta dada

que pasa por p . En efecto, esta condición se expresa algebraicamente por la ecuación

$$\vec{n} \cdot (x - p_1, y - p_2, z - p_3) = 0$$

donde $\vec{n} = (a, b, c)$ es un vector dado. Desarrollando este producto escalar, esa ecuación toma la forma

$$ax + by + cz + d = 0$$

con $d = -\vec{n} \cdot p = -(ap_1 + bp_2 + cp_3)$, o sea, se trata de la ecuación implícita de un plano.

Dos planos

$$ax + by + cz + d = 0$$

$$a'x + b'y + c'z + d' = 0$$

serán paralelos si los vectores (a, b, c) y (a', b', c') son uno múltiplo escalar del otro, y serán *ortogonales* si (a, b, c) y (a', b', c') son ortogonales.

Por ejemplo, el plano de ecuación $z = 1$ es el conjunto de puntos cuya tercera coordenada vale 1; está caracterizado como el plano que contiene, por ejemplo, al punto $(0, 0, 1)$ y que es perpendicular al eje z , o sea, al vector $(0, 0, 1)$:

$$(0, 0, 1) \cdot (x - 0, y - 0, z - 1) = 0$$

lo que da $z = 1$. Es paralelo al plano coordenado xy ($z = 0$) y está a una distancia de éste igual a 1. Es la gráfica de la función constante $f(x, y) \equiv 1$, que hace corresponder a todo $(x, y) \in \mathbb{R}^2$ el valor $1 \in \mathbb{R}$ (Fig. 1.27).

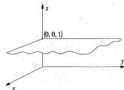


Figura 1.27

La gráfica de $z = 2x + y$ es un plano que pasa por el origen, ya que $0 = 2 \cdot 0 + 0$. Escrita la ecuación en la forma

$$2x + y - z = 0$$

vemos que el plano es perpendicular al vector $(2, 1, -1)$. Para levantar la gráfica en \mathbb{R}^3 puede ser útil también la determinación de las rectas de corte del plano con los planos coordenados: la recta intersección con el plano xy ($z = 0$) tiene por ecuaciones

$$\begin{cases} x + 2y = 0 \\ z = 0 \end{cases}$$

y la intersección con el plano yz ($x = 0$)

$$\begin{cases} x = 0 \\ z = y \end{cases}$$

El plano queda determinado en \mathbb{R}^3 por estas dos rectas (Fig. 1.28).

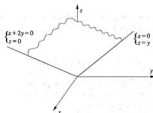


Figura 1.28

La ecuación $y = x$ representa en el plano xy una recta que pasa por el origen (Fig. 1.29). Interpretada como una ecuación en las tres variables x, y, z , su gráfica es un plano tal que, si $(x, y, 0)$ está en él, también lo está (x, y, z) para cualquier $z \in \mathbb{R}$ (no hay en la ecuación ninguna restricción en z). En consecuencia, se trata del plano que corta al plano coordenado xy según la recta $y = x$ y es paralelo al eje z (o sea, perpendicular al plano xy). Se puede tomar $p = (0, 0, 0)$ y $\vec{n} = (1, -1, 0)$; Fig. 1.29). La ecuación $y = x$ no define a z como función de (x, y) ; sí define a x como función de (y, z) y a y como función de (x, z) ; los planos "verticales", o sea, aquéllos de ecuación $ax + by + d = 0$, con $c = 0$, no son gráficas de funciones $z = f(x, y)$.

En la ecuación de una recta

$$y = mx + d \quad (1.61)$$

el número m es la pendiente de la recta; da una medida de la inclinación de la recta respecto al eje x ; es decir, una medida del "ritmo" de variación de la

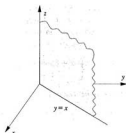


Figura 1.29

variable dependiente y respecto a la variable independiente x , o, en términos más precisos, de la variación que sufre y como respuesta a un cambio unitario de la variable x :

$$y(x_0 + 1) - y(x_0) = m(x_0 + 1) + d - [mx_0 + d] = m$$

Vemos que ese "ritmo" es constante, independiente del valor x_0 considerado (m es, como sabemos, la derivada de f en cualquier punto $x_0 \in \mathbb{R}$).

Para una función de dos variables

$$z = f(x, y) = ax + by + c \quad (1.62)$$

y el plano que constituye su representación gráfica se pueden hacer consideraciones análogas. Para empezar, es útil estudiar la dependencia de z respecto a cada una de las variables x e y separadamente, suponiendo fija la otra. Si, dado un punto $(x_0, y_0) \in \mathbb{R}^2$, mantenemos y_0 constante e incrementamos x_0 en una cantidad h , la variación de f será

$$f(x_0 + h, y_0) - f(x_0, y_0) = a(x_0 + h) + by_0 + c - [ax_0 + by_0 + c] = ah$$

Es decir, a es la variación que sufre z como consecuencia de un cambio unitario $h = 1$, en la variable x (y esto, cualquiera que sea el punto $(x_0, y_0) \in \mathbb{R}^2$). Es la pendiente de la recta sobre el plano (1.62) que resulta al cortar éste por el plano vertical $y = y_0$ (Fig. 1.30); resulta natural llamarla pendiente del plano $z = ax + by + c$ en la dirección del eje x .

Análogamente, b es la pendiente del plano en la dirección del eje y (Fig. 1.31). (a y b serán las derivadas parciales de la función $f(x, y)$ respecto a x e y , respectivamente.)

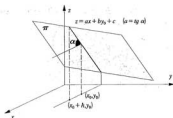


Figura 1.30

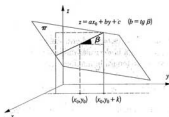


Figura 1.31

De manera general, la variación que experimenta z cuando nos desplazamos desde (x_0, y_0) en la dirección dada por un vector $\vec{u} = (u_1, u_2)$, unitario (o sea, de longitud unidad: $u_1^2 + u_2^2 = 1$) vendrá dada por

$$\begin{aligned} f(x_0 + \lambda u_1, y_0 + \lambda u_2) - f(x_0, y_0) &= \\ a(x_0 + \lambda u_1) + b(y_0 + \lambda u_2) + c - ax_0 - by_0 - c &= \lambda(au_1 + bu_2) \end{aligned}$$

Si el desplazamiento es de magnitud 1, o sea, $\lambda = 1$, se tendrá

$$f(x_0 + u_1, y_0 + u_2) - f(x_0, y_0) = au_1 + bu_2 = (a, b) \cdot (u_1, u_2) \quad (1.63)$$

que es la pendiente de la recta sobre el plano (1.62) que resulta al cortar éste por el plano vertical

$$\begin{cases} x = x_0 + \lambda u_1 \\ y = y_0 + \lambda u_2 \\ z = \mu \end{cases} \quad \circ \quad u_2(x - x_0) - u_1(y - y_0) = 0$$

que, a su vez, corta al plano coordenado $z = 0$ según la recta que pasa por (x_0, y_0) y contiene a $u = (u_1, u_2)$ (Fig. 1.32).

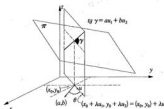


Figura 1.32

Diremos que $au_1 + bu_2$ es la pendiente del plano en la dirección dada por el vector $u = (u_1, u_2)$ (será la derivada direccional de $f(x, y)$ según el vector u). ¿En qué dirección se tendrá la máxima pendiente? De (1.63) se tiene:

$$au_1 + bu_2 = (a, b) \cdot (u_1, u_2) = |(a, b)| |u| \cos \theta = \sqrt{a^2 + b^2} \cdot \cos \theta \quad (1.64)$$

siendo θ el ángulo que forman los vectores \vec{u} y (a, b) en el plano xy .

Es claro que el máximo valor de (1.64) se obtiene para $\theta = 0$ ($\cos \theta = 1$), es decir, en la dirección del vector (a, b) , el cual recibe el nombre de *gradiente* de la función $z = ax + by + c$ (o del plano que constituye su gráfica). Se denota por $\text{grad } f$ (EJERCICIO: Compruébese que la ecuación del plano que pasa por un punto dado $p = (p_1, p_2, p_3)$ y tiene un gradiente (a, b) dado es

$$(a, b, -1) \cdot (x - p_1, y - p_2, z - p_3) = 0$$

Gráficas de funciones no lineales La gráfica de una función lineal de una variable $y = mx + h$ es una línea recta y la de una función no lineal $y = f(x)$ es una línea curva. De manera análoga, la gráfica de una función lineal de dos variables es una superficie plana y la de una función no lineal una superficie "curva". Con la idea de captar este hecho conviene repasar, como ya dijimos, algunas de las superficies no planas del espacio más comunes e importantes.

Esferas El conjunto de todos los puntos (x, y, z) del espacio que están a una distancia fija r de un punto dado (x_0, y_0, z_0) es una *esfera* de radio r y centro (x_0, y_0, z_0) (Fig. 1.33).

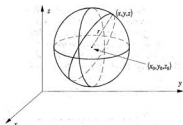


Figura 1.33

La ecuación algebraica que satisfacen las coordenadas x, y, z de los puntos de esa esfera es, por tanto,

$$\sqrt{(x-x_0)^2 + (y-y_0)^2 + (z-z_0)^2} = r$$

o, equivalentemente,

$$(x-x_0)^2 + (y-y_0)^2 + (z-z_0)^2 = r^2 \quad (1.65)$$

(Reservamos aquí el nombre de *esfera* para la superficie —esférica—; a la “esfera sólida”, cuyos puntos satisfacen $(x-x_0)^2 + (y-y_0)^2 + (z-z_0)^2 \leq r^2$ la denominaremos *bola*).

La *esfera unidad* es la que tiene por centro el origen $(0, 0, 0)$ y radio 1. Su ecuación es

$$x^2 + y^2 + z^2 = 1 \quad (1.66)$$

Resolviendo (1.66) en z se tienen dos valores

$$z = \pm \sqrt{1 - x^2 - y^2}$$

para cada par (x, y) tal que $x^2 + y^2 \leq 1$. Así pues, la ecuación (1.66) define dos funciones $z = f(x, y)$ de dos variables. La gráfica de la primera de ellas, $z = \sqrt{1 - x^2 - y^2}$, es la semiesfera que está sobre el plano xy y la de la segunda, la semiesfera inferior (Fig. 1.3).

Cilindros La ecuación $x^2 + z^2 = r^2$ representa en el plano xz la circunferencia de centro el origen y radio r . Si interpretamos la ecuación $x^2 + z^2 = r^2$

como una ecuación en las tres variables x, y, z , vemos que no hay ninguna restricción en y ; por ejemplo, $(3, 0, 4)$ pertenece a la gráfica de $x^2 + z^2 = 25$, y lo mismo $(3, y, 4)$ para cualquier y : el corte de dicha gráfica con cada plano $y = \text{constante}$ es la traslación paralelamente al eje y de la circunferencia $x^2 + z^2 = 25$ del plano xz . La gráfica en \mathbb{R}^3 de $x^2 + z^2 = r^2$ es un cilindro circular recto (Fig. 1.34). La circunferencia $x^2 + z^2 = r^2$ del plano xz es

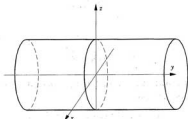


Figura 1.34

la curva directriz del cilindro y las rectas paralelas al eje y que pasan por dicha circunferencia, las generatrices. La ecuación $x^2 + z^2 = r^2$ define, como antes, dos funciones $z = \pm\sqrt{r^2 - x^2}$ en las que no aparece explícitamente la variable y . La gráfica de $z = \sqrt{r^2 - x^2}$ es el semicilindro que está por encima del plano xy , y la de $z = -\sqrt{r^2 - x^2}$, el que está por debajo (la ecuación del cilindro "vertical" $x^2 + y^2 = r^2$ no define a z como función de (x, y) ; sí define a x como función de (y, z) y también a y como función de (x, z)).

En general, dada una curva C en un plano y una recta l no contenida en ese plano, el conjunto de todas las rectas paralelas a l que cortan a C se dice que es un cilindro; C es la curva generatriz del cilindro y esas rectas paralelas son las generatrices. El cilindro se dice que es recto si sus generatrices son perpendiculares al plano que contiene a C . (EJERCICIO: Esbozar la gráfica en \mathbb{R}^3 del cilindro $z = y^2$. Obsérvese que la gráfica en el espacio de una ecuación en dos de las tres variables x, y, z es un cilindro cuyas generatrices son paralelas al eje de la variable que falta).

Elipsoides Si la esfera unidad se "estira" en la dirección del eje x por un factor a , en la dirección del eje y por un factor b y en la del eje z por un factor c , el punto (x, y, z) se transforma en el punto (ax, by, cz) y la esfera en la superficie de ecuación

$$\left(\frac{x}{a}\right)^2 + \left(\frac{y}{b}\right)^2 + \left(\frac{z}{c}\right)^2 = 1 \quad (1.67)$$

la cual recibe el nombre de *elipsoide* de semiejes a, b y c (Fig. 1.35). Las intersecciones con los planos coordenados (y con planos paralelos a éstos) son *elipses*. Si, en particular, $a = b = c$, se tiene la esfera de radio a y centro el origen.

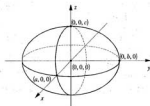


Figura 1.35

La ecuación (1.67) define también *dos* funciones:

$$z = \pm c \sqrt{1 - \frac{x^2}{a^2} - \frac{y^2}{b^2}} \quad (1.68)$$

y el elipsoide es la unión de las gráficas de ambas funciones.

Hiperboloides La superficie de ecuación

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = 1 \quad (1.69)$$

es un *hiperboloide de una hoja* (y lo mismo las superficies de ecuaciones $-\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1$, $\frac{x^2}{a^2} - \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1$). Para representarla, comprobamos que las intersecciones con planos horizontales $z = k$ son *elipses*:

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1 + \frac{k^2}{c^2}$$

mientras que la intersección con el plano yz (y también con los planos paralelos a él $x = k$) es una *hipérbola*

$$\frac{y^2}{b^2} - \frac{z^2}{c^2} = 1$$

(y análogamente para el plano xz ; Fig 1.36).

La superficie de ecuación

$$\frac{z^2}{c^2} - \frac{x^2}{a^2} - \frac{y^2}{b^2} = 1 \quad (1.70)$$

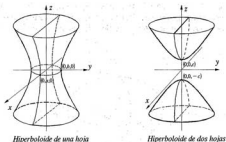


Figura 1.36

es un *hiperboloide de dos hojas* (y lo mismo las superficies de ecuaciones $\frac{x^2}{a^2} - \frac{y^2}{b^2} - \frac{z^2}{c^2} = 1$, $-\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = 1$). Las intersecciones con planos horizontales $z = k$, con $|k| > c$, son elipses mientras que las intersecciones con los planos xz e yz (y sus paralelos $y = k$ y $x = k$) son hipérbolas (Fig. 1.36).

Las ecuaciones (1.69) y (1.70) definen cada una de ellas *dos* funciones $z = f(x, y)$ tales que la unión de sus gráficas componen el correspondiente hiperboloide.

Paraboloides La gráfica de la función

$$z = \frac{x^2}{a^2} + \frac{y^2}{b^2} \quad (1.71)$$

es una superficie que se llama *paraboloide elíptico*. Los cortes con planos horizontales $z = k$, $k > 0$, son elipses (en particular, circunferencias si $a = b$) mientras que los cortes con planos $x = k$ e $y = k$ son parábolas (Fig. 1.37).

La gráfica de la función

$$z = \frac{y^2}{b^2} - \frac{x^2}{a^2} \quad (1.72)$$

es un *paraboloide hiperbólico* y su aspecto es el de una *silla de montar* o un *puerto de montaña*. Sus intersecciones con planos horizontales $z = k$ son hipérbolas (un par de rectas que se cortan en el origen si $k = 0$), y con los planos $x = k$ e $y = k$, parábolas (Fig. 1.38).

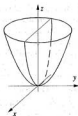


Figura 1.37

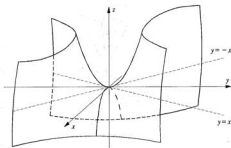


Figura 1.38

De manera general, una superficie cuádrica en el espacio es la gráfica de una ecuación algebraica de segundo grado:

$$Ax^2 + By^2 + Cz^2 + Dxy + Exz + Fyz + Gx + Hy + Iz + J = 0 \quad (1.73)$$

o sea, el conjunto de puntos del espacio cuyas coordenadas (x, y, z) satisfacen (1.73).

Las cuádricas son la contrapartida tridimensional de las secciones cónicas en el plano, que son las gráficas de las ecuaciones de segundo grado en dos variables

$$Ax^2 + Bxy + Cy^2 + Dx + Ey + F = 0 \quad (1.74)$$

Vemos como la intersección de una cuádrica con cada uno de los planos coordenados (y sus planos paralelos) es una cónica.

Así como existen tres tipos básicos de cónicas: elipse, hipérbola y parábola, cuyas ecuaciones canónicas o reducidas—previo cambio de variables—son, respectivamente, (1.33), (1.34) y (1.36), hay cinco tipos básicos de cuádricas, que son las que hemos mencionado con sus respectivas ecuaciones canónicas: elipsoide, hiperboloide de una hoja, hiperboloide de dos hojas, paraboloide elíptico y paraboloide hiperbólico. (En uno y otro caso, esos tipos básicos son las denominadas cónicas y cuádricas no degeneradas reales; las cuádricas degeneradas pueden ser conos, de ecuación canónica $\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = 0$ (Fig. 1.39), cilindros elípticos: $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$, cilindros hiperbólicos: $\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$, cilindros parabólicos: $ax^2 + y = 0$, y pares de planos, paralelos o no. Véase [8] para un estudio detallado de las cuádricas en el espacio).

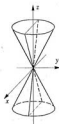


Figura 1.39

En todos los ejemplos anteriores vemos como una superficie, esfera, elipsoide, etcétera, cuya imagen nos aparece clara a la intuición sobre la base de conceptos de la geometría elemental del plano y del espacio, está constituida por uno o varios trozos cada uno de los cuales es la gráfica de una función $z = f(x, y)$ de dos variables. Este es el punto de partida para la definición abstracta de superficie en \mathbb{R}^3 y, en general, de variedad en un espacio euclídeo \mathbb{R}^n , cuestión en la que no entraremos en este libro (véase, por ejemplo, [4]). Nos aprovecharemos, eso sí, de la asociación superficie \rightarrow gráfica de $z = f(x, y)$ para dar una idea intuitiva de importantes resultados de la teoría que estableceremos analíticamente de un modo riguroso. Como decíamos en el caso de rectas y cónicas en el plano, el estudio de planos y cuádricas en el espacio puede llevarse a cabo de manera muy completa por métodos puramente algebraicos y geométricos. Pero para estudiar la gráfica de una función general de dos variables —y, en definitiva, el comportamiento de la función y que ello nos sirva para el estudio de funciones de cualquier número de variables—necesitaremos desarrollar los instrumentos análogos a los que antes recordábamos

del cálculo infinitesimal de una variable. Por ejemplo, podremos dar sentido a la pendiente de una superficie según una dirección —como hacíamos para los planos— de manera análoga a como la derivada de una función de una variable en un punto da una medida de la pendiente —inclinación, ritmo de crecimiento— de la curva en ese punto. También será una cuestión importante el dilucidar si una ecuación de tres variables $F(x, y, z) = 0$ define o no a una variable como función de las dos restantes.

Coordenadas cilíndricas y esféricas En el espacio, hay dos sistemas de coordenadas alternativos al de coordenadas rectangulares o cartesianas que son muy útiles para realizar determinados cálculos. El primero es el de *coordenadas cilíndricas*, que combina las coordenadas polares del plano con la coordenada z de las rectangulares del espacio. Las coordenadas cilíndricas del punto P del espacio, de coordenadas cartesianas (x, y, z) , son los números (r, θ, z) definidos por (figura 1.40)

$$x = r \cos \theta, \quad y = r \sin \theta, \quad z = z$$

(los puntos del eje z quedan identificados por $r = 0$ y la correspondiente

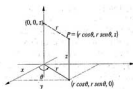


Figura 1.40

coordenada z , y no tienen asignado valor de θ). El lugar geométrico de los puntos de \mathbb{R}^3 para los que $r = a$ es un cilindro recto de radio a y eje el eje z ; de ahí el nombre de coordenadas cilíndricas.

Las *coordenadas esféricas* de un punto $P = (x, y, z)$ son los números (ρ, θ, ϕ) definidos por (figura 1.41)

$$x = \rho \sin \phi \cos \theta, \quad y = \rho \sin \phi \sin \theta, \quad z = \rho \cos \phi$$

donde

$$\rho \geq 0, \quad \theta \in [0, 2\pi), \quad \phi \in [0, \pi]$$

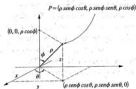


Figura 1.41

Como se ve, $\rho = \sqrt{x^2 + y^2 + z^2}$ es la distancia del punto al origen de coordenadas cartesianas, θ (la *longitud*) es el mismo ángulo que en las cilíndricas y ϕ (la *colatitud*) es el ángulo que forman el semieje z positivo y el segmento que une el punto con el origen. La ecuación en coordenadas esféricas de una esfera con centro en el origen y radio a es muy simple: $\rho = a$.

1.3 Curvas de nivel

Las gráficas en el espacio tridimensional \mathbb{R}^3 son con frecuencia difíciles de interpretar, sobre todo cuando tratemos con funciones generales $f(x, y)$ no necesariamente lineales. Hay otro modo de disponer de una imagen gráfica del comportamiento de $z = f(x, y)$ como función de x e y . Consiste en dibujar conjuntos en el plano xy para los que la función toma determinados valores fijos (por ejemplo, regularmente espaciados: $k = \dots, -2, -1, 0, 1, 2, \dots$).

El *conjunto de nivel* k de una función $f(x, y)$ es el conjunto

$$\{(x, y) \in \text{Dom } f; f(x, y) = k\}$$

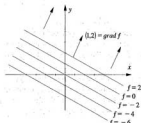
o sea, el conjunto de puntos del dominio de f a los cuales f asigna el mismo valor k (k ha de estar en el recorrido de f).

Para una función lineal $z = ax + by + c$ los conjuntos de nivel son las *líneas rectas*

$$ax + by + c = k, \quad k \in \mathbb{R}$$

Según hemos recordado en el apartado anterior, el vector (a, b) es ortogonal a la recta $ax + by + c = k$, por lo que el *gradiente de la función* $f(x, y) = ax + by + c$ es ortogonal a sus *líneas de nivel*. Por así decir, el gradiente (a, b) va señalando la dirección de (máximo) crecimiento de la función $f(x, y) = ax + by + c$ (Fig. 1.42).

Este hecho puede aprovecharse para resolver gráficamente *problemas de optimización* de funciones de dos variables, o sea, problemas de determinación

Figura 1.42 Líneas de nivel de $f(x, y) = x + 2y - 2$

de máximos y mínimos. Dada una función $f(x, y)$ y un conjunto $S \subseteq \text{Dom } f$ del plano \mathbb{R}^2 se dice que f alcanza su máximo en S en el punto $(x_0, y_0) \in S$ si $f(x, y) \leq f(x_0, y_0)$ para todo $(x, y) \in S$. $f(x_0, y_0)$ es el valor máximo de f en S , y lo representamos por

$$\max_S f(x, y)$$

Si $f(x, y) < f(x_0, y_0)$ para todo $(x, y) \in S$ distinto de (x_0, y_0) se dice que f tiene un máximo estricto en (x_0, y_0) . Se definen de manera análoga *mínimo* y *mínimo estricto* cambiando el signo de las desigualdades.

El problema

$$\begin{aligned} &\text{opt } f(x, y), (x, y) \in S \\ &(\text{o bien } \text{opt}_S f(x, y)) \end{aligned}$$

(que se lee "optimizar la función $f(x, y)$ en el conjunto $S \subseteq \mathbb{R}^2$ ", donde el término optimizar vale indistintamente para maximizar o minimizar, según de lo que se trate) consiste en determinar el punto (o puntos) de S en que la función f alcanza su valor óptimo (máximo o mínimo).

Supongamos que se tiene planteado el problema

$$\min_S f(x, y)$$

(minimizar f en el conjunto S) para una función lineal $f(x, y) = ax + by + c$ y que se tiene

$$\min_S f(x, y) = m$$

es decir,

- (i) $f(x, y) \geq m$ para todo $(x, y) \in S$,
- (ii) $f(x_0, y_0) = m$ para algún $(x_0, y_0) \in S$

La recta de nivel $ax + by + c = m$ divide al plano en dos semiplanos: el semiplano positivo $\{(x, y); f(x, y) \geq m\}$ y el negativo $\{(x, y); f(x, y) \leq m\}$ (véase la Fig. 1.43). La condición $f(x, y) \geq m$ para todo $(x, y) \in S$ quiere decir que S está en el semiplano positivo de $f(x, y) = m$ (hacia el que apunta $\text{grad } f$), y la condición (ii) que al menos un punto (x_0, y_0) de S está sobre la propia línea de nivel. Una línea de nivel con estas dos propiedades se dice que es una recta soporte de S ortogonal a $\text{grad } f$.

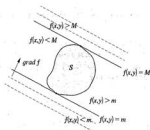


Figura 1.43 Resolución gráfica de problemas de optimización

El problema de minimizar $f(x, y)$ en S se reduce por tanto al de determinar tal recta soporte. En situaciones no muy complicadas, bastará para ello dibujar con suficiente precisión S y $\text{grad } f$ (la recta $f(x, y) = m$ será la “primera” línea de nivel que “toque” S . A su vez, la “última” línea de nivel que lo haga, $f(x, y) = M$, proporciona el valor máximo M de $f(x, y)$ en S ; los problemas de minimizar y/o maximizar $f(x, y)$ en S no tendrán solución si no hay tales líneas de nivel “primera” y/o “última”).

EJEMPLO 1. Consideremos la función $f(x, y) = 2x + y + 5$ y el círculo $S = \{(x, y); x^2 + y^2 \leq 1\}$. Por geometría elemental, las rectas soporte de S han de ser tangentes a la circunferencia $x^2 + y^2 = 1$. Las que son ortogonales a $\text{grad } f = (2, 1)$, o sea, las líneas de nivel $2x + y + 5 = k$ tangentes a dicha circunferencia, estarán dadas por los valores de k tales que el sistema

$$\begin{cases} 2x + y + 5 = k \\ x^2 + y^2 = 1 \end{cases}$$

que da los puntos comunes a la recta $2x + y + 5 = k$ y la circunferencia $x^2 + y^2 = 1$ — tenga una única solución. Imponiendo en su resolución esta condición (el discriminante de la ecuación de segundo grado que resulta ha de ser nulo) se obtiene

$$-(k-5)^2 + 5 = 0$$

es decir, $k_1 = 5 - \sqrt{5}$ y $k_2 = 5 + \sqrt{5}$. En consecuencia,

$$\min_S f(x, y) = 5 - \sqrt{5}, \quad \max_S f(x, y) = 5 + \sqrt{5} \quad (1.75)$$

El mínimo se alcanza en el punto $(-\frac{2\sqrt{5}}{5}, -\frac{\sqrt{5}}{5})$ y el máximo en $(\frac{2\sqrt{5}}{5}, \frac{\sqrt{5}}{5})$ (Fig. 1.44; más adelante, en el apartado 6.1, veremos que una función lineal no puede alcanzar sus valores máximo y mínimo en el "interior" de S).

En este caso tan simple se puede dar una demostración analítica de estos resultados completamente rigurosa: se tiene

$$2x + y = (2, 1) \cdot (x, y) = |(2, 1)| |(x, y)| \cos \theta = \sqrt{5} \sqrt{x^2 + y^2} \cos \theta \quad (1.76)$$

siendo θ el ángulo que forman los vectores $\text{grad } f = (2, 1)$ y (x, y) . Con la restricción $\sqrt{x^2 + y^2} \leq 1$, el valor máximo de (1.76) se obtiene para $x^2 + y^2 = 1$ y $\cos \theta = 1$ y el mínimo para $x^2 + y^2 = 1$ y $\cos \theta = -1$, con lo que valen, respectivamente, $\sqrt{5}$ y $-\sqrt{5}$ y se tiene el resultado (1.75).

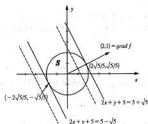


Figura 1.44

Si el problema fuese optimizar $f(x, y) = 2x + y + 5$ en el círculo abierto $S_0 = \{(x, y); x^2 + y^2 < 1\}$ (o sea, sin la circunferencia $x^2 + y^2 = 1$), obtendríamos que, desde luego,

$$\begin{aligned} f(x, y) &< 5 + \sqrt{5} && \text{para todo } (x, y) \in S_0 \\ f(x, y) &> 5 - \sqrt{5} && \text{para todo } (x, y) \in S_0 \end{aligned}$$

pero no hay ningún punto de S_0 en el que f valga exactamente $5 + \sqrt{5}$ ni tampoco $5 - \sqrt{5}$. Los problemas

$$\max_{S_0} f(x, y), \quad \min_{S_0} f(x, y) \quad (1.77)$$

no tienen, pues, solución. $M = 5 + \sqrt{5}$ es el supremo ($\sup f = 5 + \sqrt{5}$) y $m = 5 - \sqrt{5}$ es el ínfimo ($\inf f = 5 - \sqrt{5}$) de la función f en el conjunto $S_0 = \{(x, y); x^2 + y^2 < 1\}$. (Recuérdese que $M \in \mathbb{R}$ ($m \in \mathbb{R}$) es el supremo (ínfimo) de un conjunto A de números reales si $r \leq M$ ($r \geq m$) para todo $r \in A$ y cualquier otro número que verifique esta propiedad es mayor que M (menor que m). Si no existe número real M (m) tal que $r \leq M$ ($r \geq m$) para todo $r \in A$ se dice que $\sup A = \infty$ ($\inf A = -\infty$). Cuando $\sup A = M$ ($\inf A = m$) pertenece al propio A se dice que M es el máximo (mínimo) de A). En los dos problemas de optimización de este ejemplo, $\sup f(x, y)$ e $\inf f(x, y)$ son ambos finitos, pero en el primer caso $\sup f(x, y)$ e $\inf f(x, y)$ se alcanzan en S —o sea, hay puntos en el propio S en los que la función toma esos valores, y por eso son, respectivamente, $\max_S f(x, y)$ y $\min_S f(x, y)$ — mientras que en el segundo los valores $\sup_{S_0} f(x, y)$ e $\inf_{S_0} f(x, y)$ no se alcanzan en S_0 y por ello los problemas (1.77) no tienen solución.)

EJEMPLO 2. Optimizar la función $f(x, y) = 2x + y + 5$ en el conjunto S de puntos (x, y) del plano determinado por las desigualdades

$$\begin{aligned} 3x + y &\geq 4 & 4x - y &\leq 14 \\ -x + 3y &\geq 2 & -2x + 5y &\leq 20 \end{aligned} \quad (1.78)$$

Las desigualdades que determinan S responden todas a la forma

$$g_i(x, y) \leq c_i \quad (1.79)$$

donde $g_i(x, y) = a_i x + b_i y$ es una función lineal (en las dos primeras, basta cambiar de signo). $g_i(x, y) \leq c_i$ define el semiplano negativo determinado por la recta $g_i(x, y) = c_i$; la región S es el polígono intersección de todos los semiplanos definidos en (1.78); dibujándolo, vemos fácilmente que el mínimo de f se alcanza en el vértice $(1, 1)$ y el máximo en el también vértice $(5, 6)$ (Fig. 1.45).

Un programa lineal consiste en optimizar una función lineal sobre un conjunto S que está definido mediante desigualdades que involucran funciones

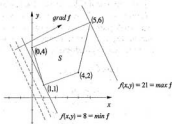


Figura 1.45

también lineales:

$$\begin{aligned} \text{opt } f(x, y) &= ax + by + c \\ \text{s. a. (sujeta a)} \quad &g_i(x, y) = a_i x + b_i y \leq c_i, \quad i = 1, \dots, r \end{aligned} \quad (1.80)$$

(las desigualdades $a_i x + b_i y \leq c_i$ se dice que son *restricciones de desigualdad*).

S es una región poligonal, intersección de semiplanos, y se puede demostrar analíticamente que, de existir, los óptimos se alcanzan siempre en vértices de S , o sea, puntos en los que se satisfacen dos de las ecuaciones $g_i(x, y) = c_i$.

Mediante la *programación lineal* se resuelven problemas prácticos muy interesantes, en particular en Economía cuando, por ejemplo, se plantea el objetivo de minimizar costes o maximizar beneficios en un proceso de producción.

EJEMPLO 3. Una empresa produce dos tipos de productos; en cada jornada de ocho horas, produce x_1 unidades del tipo I y x_2 del tipo II. Si el beneficio neto que obtiene por la venta de una unidad de cada producto es, respectivamente, de 4 y 3 unidades monetarias (u.m.), el beneficio total vendrá dado por la función $f(x_1, x_2) = 4x_1 + 3x_2$. Claramente, la empresa obtendría, en principio, el máximo beneficio dedicando toda su capacidad de producción a fabricar el producto I. Pero en la práctica, el proceso de producción estará sometido a *restricciones* causadas por factores como la disponibilidad de materias primas, mano de obra o de tiempo de uso de la maquinaria. Si, por ejemplo, el producto I requiere cierto material del que no se dispone cantidad más que para producir 10 unidades en cada jornada, habrá de ser $x_1 \leq 10$; si el de tipo II requiere otro material que limita su producción a un máximo de 12 unidades por jornada, habrá de cumplirse $x_2 \leq 12$. Si, por otro lado, se emplean, respectivamente, 40 y 20 minutos en fabricar una unidad de cada uno de los productos I y II,

se tendrá $\frac{2}{3}x_1 + \frac{1}{3}x_2 \leq 8$. Finalmente, hay otras dos restricciones debidas a la naturaleza del problema: ha de ser $x_1 \geq 0$, $x_2 \geq 0$ pues no tiene sentido fabricar un número negativo de piezas. Con todas esas restricciones, se trata de determinar la combinación (x_1, x_2) de productos que proporciona el máximo beneficio. El programa lineal correspondiente es

$$\begin{aligned} &\max \{4x_1 + 3x_2\} \\ &\text{s.a.} \\ &\quad x_1 \leq 10 \\ &\quad x_2 \leq 12 \\ &\quad 2x_1 + x_2 \leq 24 \\ &\quad x_1, x_2 \geq 0 \end{aligned} \tag{1.81}$$

La función $f(x_1, x_2) = 4x_1 + 3x_2$ se dice que es la *función objetivo* del programa; las restricciones definen un conjunto S del plano que se denomina *conjunto de soluciones factibles* o simplemente *conjunto factible* del programa. La resolución gráfica nos indica que la solución se alcanza en el vértice $(6, 12)$, o sea, que el beneficio máximo se obtiene fabricando 6 unidades del producto I y 12 del producto II y vale 60 u.m. (Figura 1.46).

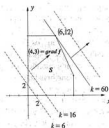
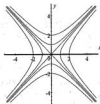


Figura 1.46

Ya se comprende que la resolución gráfica de un problema de optimización está limitada a funciones de dos variables; para funciones de tres variables ya serían necesarios dibujos en el espacio tridimensional, difíciles de realizar e interpretar, y para funciones de más de tres variables el procedimiento sería simplemente imposible. Como cabe suponer, existen métodos analíticos completamente rigurosos que permiten resolver programas lineales de cualquier número de variables y cualquier número de restricciones. Partiendo, en particular, del hecho de que hay que buscar el óptimo en los vértices de la región

factible S , se han desarrollado métodos algorítmicos de resolución apropiados para su realización en computadoras. Entre ellos destaca el método del *simplex* (véase, por ejemplo, [1]).

Curvas de nivel de funciones no lineales Hablando en general, los conjuntos de nivel de una función no lineal $z = f(x, y)$ de dos variables "usual" son curvas ("curvas de nivel"). Por ejemplo, los conjuntos de nivel de la función $z = x^2 + y^2$ son las circunferencias $x^2 + y^2 = k$, $k \geq 0$. Son las proyecciones sobre el plano xy de las intersecciones de los planos $z = k$ con el paraboloide de ecuación $z = x^2 + y^2$. El conjunto de nivel 0 se reduce al origen (Fig. 1.47). Las curvas de nivel de la función $z = y^2 - x^2$ son hipérbolas que se representan en la figura 1.48 (el conjunto de nivel 0 está formado por las rectas $y = x$ e $y = -x$).

Figura 1.47. $x^2 + y^2 = k$ Figura 1.48. $y^2 - x^2 = k$

Una representación gráfica de función de dos variables mediante sus curvas de nivel que ya resulta familiar es la que aparece en los mapas de predicción del tiempo: en ellos se dibujan las *isobaras* ("igual presión"), que son las curvas de nivel de la función presión atmosférica sobre los puntos (x, y) de un cierto territorio (que, para simplificar, se considera plano). A veces se dibujan también *isotermas* que unen puntos en los que la temperatura máxima prevista es la misma. Aunque menos populares, también son bien conocidos los mapas topográficos que describen el relieve de un territorio mediante las curvas de nivel, correspondientes a valores regularmente espaciados, de la función altura sobre el nivel del mar de cada punto. Un montañero experimentado lee perfectamente las características de dicho relieve —hecho tridimensional— en la disposición de las curvas de nivel en el mapa —instrumento bidimensional; por ejemplo, una zona en la que dichas curvas estén muy próximas indica una gran pendiente (¿por qué?); en el entorno de un puerto de montaña, las curvas de nivel presentarán un aspecto semejante, no necesariamente tan regular, al de la figura 1.47 (Fig. 1.49).



Figura 1.49

Funciones de Cobb-Douglas En Economía, es común tomar como descripción del nivel de producción Q de un determinado proceso una función de la forma

$$Q = f(x, y) = cx^r y^s \quad (1.82)$$

donde x e y denotan, respectivamente, las unidades (≥ 0) de trabajo y de capital empleadas y c , r y s son números positivos. La función (1.82) tiene la propiedad de que un decrecimiento de una de las variables (o factores de producción) puede ser compensado con un crecimiento de la otra para mantener el nivel de producción; satisface también $f(0, y) = f(x, 0) = 0$, lo que corresponde al hecho de que si no hay trabajo o capital no puede haber producción. Se suele suponer también que es una función homogénea de primer grado, o sea, que satisface

$$f(\alpha x, \alpha y) = \alpha f(x, y), \quad \alpha > 0 \quad (1.83)$$

lo que refleja, en terminología económica, *rendimientos constantes a escala*: si se duplican a la vez mano de obra y bienes de capital, se duplica el nivel de producción; no se producen economías ni deseconomías en función de la dimensión (o escala) de la producción. La corriente de la teoría económica llamada "neoclásica" considera que ésta puede ser una buena hipótesis para describir procesos de acumulación incipiente de capital sin cambio de tecnología, en los que no hay saturación y los incrementos en capital y mano de obra siempre tienen "espacio libre" para situarse y producir bienes con los mismos mecanismos de producción que el resto. Suele decirse que esta situación se da con bastante aproximación en economías agrarias con muy poca población y suficiente tierra para dar trabajo a las sucesivas generaciones. Esa hipótesis de

homogeneidad implica que $r + s = 1$ (compruébese como ejercicio), con lo que la función de producción toma la forma

$$Q = f(x, y) = cx^ry^{1-r}, \quad 0 < r < 1 \quad (1.84)$$

Para cada elección de r se tendrá una función de Cobb-Douglas, como se conoce a este tipo de funciones de producción. Las curvas de nivel de (1.84) están dadas por $cx^ry^{1-r} = k$, o sea

$$y^{1-r} = \frac{k}{c} \frac{1}{x^r} \quad (1.85)$$

que da la ecuación explícita

$$y = \frac{k_1}{x^p} \quad (1.86)$$

donde $p = \frac{1}{1-r}$ y $k_1 = \left(\frac{k}{c}\right)^{\frac{1}{1-r}}$. Nótese que $p \in (0, \infty)$ y que k_1 es una función creciente de k . Estas curvas de nivel se denominan *isocuantas* y su aspecto se muestra en la figura 1.50.

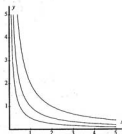


Figura 1.50 $y = \frac{k_1}{x^p}$

La pendiente de una isocuanta $y = k_1x^{-p}$ es

$$\frac{dy}{dx} = -k_1px^{-p-1} = -\frac{k_1p}{x^{p+1}} \quad (1.87)$$

La interpretación de (1.87) en términos aproximados:

$$\Delta y \simeq -\frac{k_1p}{x^{p+1}} \Delta x \quad (1.88)$$

nos da el valor (aproximado) de la variación Δy en unidades de capital que hay que realizar para mantener la misma producción cuando el trabajo se modifica en Δx unidades. Obsérvese que Δx e Δy tienen signos opuestos. La *tasa marginal de sustitución* es, por definición, la cantidad

$$-\frac{dy}{dx} = \frac{k_1 p}{x^{p+1}} > 0 \quad (1.89)$$

y viene a representar el aumento de capital que hay que realizar para mantener la producción cuando el factor trabajo disminuye en una unidad ($\Delta x = -1$ en (1.88)).

También es posible resolver gráficamente problemas de optimización sencillos para funciones no lineales. No es, desde luego, un método de gran alcance práctico, pero puede ser útil para ilustrar y motivar resultados generales, sobre todo cuando dispongamos de más instrumentos teóricos, como el *gradiente* de una función no lineal (que, a diferencia de las funciones lineales, será variable con el punto pero que también nos señalará el crecimiento de la función a lo largo de sus curvas de nivel), el teorema de Weierstrass, etcétera. Busquemos, por ejemplo, los óptimos de $f(x, y) = x^2 + y^2$ en el conjunto $S = \{(x, y) \in \mathbb{R}^2; (x-2)^2 + y^2 \leq 1\}$, que es el círculo (con la circunferencia) de centro $(2, 0)$ y radio 1 (Fig. 1.51). Las curvas de nivel de la función objetivo son las circunferencias

$$x^2 + y^2 = r^2 \quad (1.90)$$

de centro el origen y radio r . El "nivel" de f va creciendo con r ; la "primera" de las circunferencias (1.90) que toca S es claramente la de radio $r \stackrel{S}{=} 1$, y lo hace en el punto $(1, 0)$; la última es la de radio $r = 3$ y lo hace en el punto $(3, 0)$. Así pues

$$\begin{aligned} \min_S f(x, y) &= 1^2 = 1 = f(1, 0) \\ \max_S f(x, y) &= 3^2 = 9 = f(3, 0) \end{aligned}$$

Es muy fácil, como veremos en el capítulo 6, justificar este resultado analíticamente.

1.4 Funciones de n variables. El espacio \mathbb{R}^n .

La variable independiente multidimensional $x = (x_1, \dots, x_n)$ de una función de n variables se mueve en el conjunto \mathbb{R}^n de n -uplas ordenadas de números reales. En el corto estudio realizado hasta aquí de las funciones de dos variables $z = f(x, y)$, hemos utilizado conceptos y resultados de la geometría

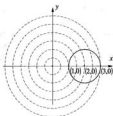


Figura 1.51

elemental del plano \mathbb{R}^2 y el espacio \mathbb{R}^3 tales como suma de vectores, paralelismo, producto escalar de vectores, ortogonalidad, distancia entre puntos, longitud de un vector, ángulo, ..., y los seguiremos utilizando en los capítulos siguientes; en particular, la *distancia*, que permite cuantificar con precisión la idea clave de *proximidad* en el concepto de continuidad que estudiaremos en el capítulo siguiente. Para llevar a cabo un estudio análogo para las funciones de un número cualquiera n de variables es necesario disponer de los correspondientes conceptos y resultados en el espacio \mathbb{R}^n ; el estudio de éstos forma parte del contenido de un primer curso típico de álgebra lineal, contenido que supondremos asimilado para poder progresar en el estudio de funciones de n variables. De todos modos, y como una simple guía de lo mínimo indispensable, repasaremos aquí los conceptos necesarios en esta primera parte (en la segunda parte se necesitará también la correspondencia entre aplicaciones lineales y matrices).

\mathbb{R}^n es, como hemos recordado, el conjunto de n -uplas ordenadas de números reales (x_1, \dots, x_n) . Nos referiremos a sus elementos como *puntos* o *vectores* por extensión de las interpretaciones geométricas en plano y el espacio que hemos repasado en el apartado 2.

Dados dos vectores (puntos) $x = (x_1, \dots, x_n)$, $y = (y_1, \dots, y_n)$ de \mathbb{R}^n , su *suma* es el vector

$$x + y = (x_1 + y_1, \dots, x_n + y_n) \quad (1.91)$$

que se obtiene sumando (en \mathbb{R}) coordenada a coordenada. En \mathbb{R}^2 , la suma se realiza gráficamente mediante la *ley del paralelogramo* (Fig. 1.4).

Si $c \in \mathbb{R}$ y $x \in \mathbb{R}^n$ se define

$$cx = (cx_1, \dots, cx_n) \quad (1.92)$$

multiplicando cada coordenada de x por el número (escalar) c .

Finalmente, escribimos en \mathbb{R}^n $0 = (0, \dots, 0)$, $-x = (-1)x = (-x_1, \dots, -x_n)$, y utilizaremos la expresión $x - y$ como abreviatura de $x + (-y)$.

Las propiedades asociativa, conmutativa y distributiva de los números reales implican, como se comprueba fácilmente, las siguientes propiedades esenciales de las operaciones de *adición de vectores* y *multiplicación por escalares* definidas en (1.91) y (1.92):

$$(EV1) \quad x + (y + z) = (x + y) + z$$

$$(EV2) \quad x + y = y + x$$

$$(EV3) \quad x + 0 = x$$

$$(EV4) \quad x + (-x) = 0$$

$$(EV5) \quad (cd)x = c(dx)$$

$$(EV6) \quad (c + d)x = cx + dx$$

$$(EV7) \quad c(x + y) = cx + cy$$

$$(EV8) \quad 1x = x$$

donde x, y, z son puntos genéricos de \mathbb{R}^n y c, d son números reales cualesquiera.

En general, un *espacio vectorial (real)* es un conjunto E dotado de dos aplicaciones $E \times E \rightarrow \mathbb{R}$ y $\mathbb{R} \times E \rightarrow E$, llamadas *adición de vectores* y *multiplicación por escalares*, que poseen las propiedades (EV1) a (EV8) ((EV3) dice que existe un elemento *neutro*, denotado por 0 , tal que $x + 0 = x$ para todo $x \in E$, y (EV4) que para cada $x \in E$ existe un elemento $-x \in E$ —su elemento *opuesto*— tal que $x + (-x) = 0$). \mathbb{R}^n es pues un espacio vectorial real (“el” espacio vectorial por antonomasia, aunque hay muchos otros espacios vectoriales importantes en matemáticas).

Una *combinación lineal* de los vectores v_1, \dots, v_m es un vector de la forma $c_1 v_1 + \dots + c_m v_m$ con $c_k \in \mathbb{R}$. Un subconjunto B de un espacio vectorial E se dice que es *linealmente dependiente* si existen elementos $v_1, \dots, v_m \in B$ y escalares c_1, \dots, c_m no todos nulos tales que

$$c_1 v_1 + \dots + c_m v_m = 0 \quad (1.93)$$

B es *linealmente independiente* si no es linealmente dependiente, o sea, si de toda ecuación de la forma (1.93) con $v_1, \dots, v_m \in B$ se deduce que $c_1 = \dots = c_m = 0$.

Se dice que un subconjunto G de vectores *genera* E si todo elemento $x \in E$ es una combinación lineal de elementos de G : $x = c_1 v_1 + \dots + c_m v_m$, con $v_1, \dots, v_m \in G$. Una *base* de E es un conjunto independiente que genera E .

E es un *espacio vectorial de dimensión finita* si posee una base formada por una cantidad finita de vectores. Se puede demostrar que *toda* base de un espacio vectorial de dimensión finita tiene el mismo número n de elementos; n

es la *dimensión* del espacio. Si $B = \{u_1, \dots, u_n\}$ es una base de E , todo vector $x \in E$ se puede escribir de modo único como una combinación lineal

$$x = c_1 u_1 + \dots + c_n u_n$$

c_1, \dots, c_n son las *coordenadas* de x respecto a la base B . Por ejemplo, los vectores

$$e_1 = (1, 0, 0, \dots, 0); e_2 = (0, 1, 0, \dots, 0) \dots e_n = (0, 0, 0, \dots, 1)$$

forman la *base canónica* de \mathbb{R}^n .

Un *subespacio vectorial* de E es un subconjunto no vacío $V \subset E$ tal que $u, v \in V$ implica $u + v \in V$ y $u \in V$ implica $cu \in V$ para todo $c \in \mathbb{R}$, o sea, V es subespacio vectorial de E si tiene el mismo estructura de espacio vectorial para las operaciones definidas en E .

Para progresar en el estudio de la geometría del espacio n -dimensional \mathbb{R}^n más allá de los aspectos puramente algebraicos recogidos en el concepto de espacio vectorial, es necesario disponer de las nociones de *longitud* de un vector de \mathbb{R}^n y de *ortogonalidad* de dos vectores, ambas directamente relacionadas. En \mathbb{R}^2 y \mathbb{R}^3 la longitud de un vector viene dada por el teorema de Pitágoras. De manera natural, inspirándonos en la verificación del teorema de Pitágoras en el plano, se define como *longitud* o *norma euclídea* de un vector (punto) $x = (x_1, \dots, x_n)$ de \mathbb{R}^n la cantidad

$$|x| = (x_1^2 + \dots + x_n^2)^{\frac{1}{2}} \quad (1.94)$$

(Podríamos empezar también, como hicimos en el repaso de \mathbb{R}^2 y \mathbb{R}^3 , definiendo la distancia entre dos puntos por la fórmula (1.111) que se da más adelante; en ese caso, la norma de un vector x sería la distancia entre x y el origen $(0, \dots, 0)$.)

Sean ahora $x = (x_1, \dots, x_n)$ e $y = (y_1, \dots, y_n)$ dos vectores independientes de \mathbb{R}^n . En el plano (subespacio de dimensión 2) que ambos determinan, serán ortogonales si forman un ángulo recto (figura 1.52), o sea, apelando de



Figura 1.52

nuevo al teorema de Pitágoras, si y sólo si

$$|x|^2 + |y|^2 = |x - y|^2 \quad (1.95)$$

Desarrollando (1.95) vemos que esa condición de ortogonalidad toma la forma

$$x_1 y_1 + \dots + x_n y_n = 0 \quad (1.96)$$

La cantidad

$$\langle x, y \rangle \equiv x \cdot y = x_1 y_1 + \dots + x_n y_n \quad (1.97)$$

se denomina *producto escalar* (o *interior*) de los vectores x e y . Así pues, dos vectores son ortogonales si y sólo si su producto escalar vale cero.

Obsérvese que

$$|x|^2 = \langle x, x \rangle \quad (1.98)$$

Se comprueban inmediatamente las siguientes propiedades esenciales del producto escalar (1.97):

$$\langle x, y \rangle = \langle y, x \rangle \quad (\text{es simétrico}) \quad (1.99)$$

$$\langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle; \quad \langle cx, y \rangle = c \langle x, y \rangle, \quad c \in \mathbb{R} \quad (\text{es bilineal}) \quad (1.100)$$

$$\langle x, x \rangle \geq 0 \quad \text{y} \quad \langle x, x \rangle = 0 \quad \text{si y sólo si} \quad x = 0 \quad (\text{es definido positivo}) \quad (1.101)$$

(En general, toda aplicación $E \times E \rightarrow \mathbb{R}$ que satisfaga (1.99)(1.100)(1.101), siendo E un espacio vectorial real, se dice que es un producto escalar o interior en E . En un espacio vectorial con producto escalar se dice que dos vectores son ortogonales si su producto escalar es nulo.)

Un resultado importante es la *desigualdad de Cauchy-Schwarz*:

$$|\langle x, y \rangle| \leq |x| |y| \quad (1.102)$$

Para $y = 0$, ambos miembros de (1.102) valen 0; supuesto, entonces, $y \neq 0$, se tiene para todo escalar λ

$$\begin{aligned} 0 &\leq |x + \lambda y|^2 = \langle x + \lambda y, x + \lambda y \rangle = \langle x, x \rangle + 2\lambda \langle x, y \rangle + \lambda^2 \langle y, y \rangle = \\ &= |x|^2 + 2\lambda \langle x, y \rangle + \lambda^2 |y|^2 \end{aligned} \quad (1.103)$$

por las propiedades (1.99)(1.100)(1.101). Para x, y fijados, la expresión de la derecha de (1.103) es un polinomio en λ que alcanza su mínimo global para el valor

$$\lambda_0 = -\frac{\langle x, y \rangle}{|y|^2} \quad (1.104)$$

Haciendo $\lambda = \lambda_0$ en (1.103) se tiene

$$0 \leq |x + \lambda_0 y|^2 = |x|^2 - \frac{|\langle x, y \rangle|^2}{|y|^2} \quad (1.105)$$

es decir

$$|\langle x, y \rangle|^2 \leq |x|^2 |y|^2$$

que es equivalente a (1.102). Obsérvese que la igualdad en (1.102) es equivalente a que $|x + \lambda_0 y| = 0$, o sea, a que $x + \lambda_0 y = 0$; así pues, si $y \neq 0$, $|\langle x, y \rangle| = |x| |y|$ si y sólo si x es un múltiplo escalar de y ; y $\langle x, y \rangle = |x| |y|$ si y sólo si x es un múltiplo escalar no negativo de y .

EJERCICIO. Demostrar que $\sum_{i=1}^n |x_i| \leq \sqrt{n} |x|$ para cualquier punto $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ (Indicación: utilícese la desigualdad de Cauchy-Schwarz para los vectores $(|x_1|, \dots, |x_n|)$ y $(1, \dots, 1)$).

Según recordábamos en el apartado 2, en \mathbb{R}^2 y \mathbb{R}^3 se tiene, por trigonometría elemental, que

$$\cos \theta = \frac{\langle x, y \rangle}{|x| |y|} \quad (1.106)$$

siendo θ el ángulo que forman los vectores x e y . Se toma entonces la relación (1.106) como *definición* del (coseno del) ángulo $\theta \in [0, \pi]$ que forman dos vectores de \mathbb{R}^n ; $\theta = \frac{\pi}{2}$ si y sólo si los vectores son ortogonales. Nótese que la desigualdad de Cauchy-Schwarz es equivalente a decir que $|\cos \theta| \leq 1$. Puesto que

$$|x - y|^2 = \langle x - y, x - y \rangle = |x|^2 + |y|^2 - 2\langle x, y \rangle$$

se tendrá

$$|x - y|^2 = |x|^2 + |y|^2 - 2|x||y|\cos \theta \quad (1.107)$$

que es la fórmula del coseno de la trigonometría (Fig. 1.53).

A partir de (1.99)(1.100)(1.101)(1.102) se obtienen las propiedades esenciales de la norma euclídea:

$$|x| \geq 0 \text{ para todo } x \in \mathbb{R}^n \text{ y } |x| = 0 \text{ si y sólo si } x = 0 \quad (1.108)$$

$$|x + y| \leq |x| + |y| \text{ (desigualdad triangular)} \quad (1.109)$$



Figura 1.53

$$|cx| = |c| |x|, \quad c \in \mathbb{R} \quad (1.110)$$

(Para la desigualdad triangular se parte de

$$|x+y|^2 = \langle x+y, x+y \rangle = |x|^2 + |y|^2 + 2\langle x, y \rangle$$

y se utiliza la desigualdad de Cauchy-Schwarz. En general, siendo E un espacio vectorial, toda aplicación $N : E \rightarrow [0, \infty)$ que satisfaga (1.108), (1.109) y (1.110) se dice que es una *norma* sobre E y, dotado de una norma, se dice que E es un *espacio normado*.)

La noción de longitud o norma euclídea de un vector lleva de manera natural a la noción de *distancia* entre dos puntos de \mathbb{R}^n (Fig. 1.53):

$$d(x, y) = |x - y| = \sqrt{(x_1 - y_1)^2 + \dots + (x_n - y_n)^2} \quad (1.111)$$

A partir de (1.108)(1.109)(1.110) se obtienen las propiedades esenciales de la distancia:

$$d(x, y) \geq 0 \quad \text{y} \quad d(x, y) = 0 \quad \text{si y sólo si} \quad x = y \quad (1.112)$$

$$d(x, y) = d(y, x) \quad (1.113)$$

$$d(x, z) \leq d(x, y) + d(y, z) \quad (\text{desigualdad triangular}) \quad (1.114)$$

(Véase la figura 1.54 para la justificación del término “desigualdad triangular”. En general, si E es un conjunto dotado de una aplicación $d : E \times E \rightarrow [0, \infty)$ que satisface (1.112)(1.113)(1.114) se dice que es un *espacio métrico*; la aplicación d se denomina *distancia* o *métrica*. Nótese: \mathbb{R}^n es un espacio métrico, con métrica definida por una norma, y ésta proviniendo de un producto escalar.)

La distancia (1.111) generaliza a cualquier \mathbb{R}^n la distancia entre dos puntos de la recta \mathbb{R} y permite, sobre la base de la idea de *proximidad*, desarrollar, como en \mathbb{R} , las nociones de convergencia, entorno de un punto, continuidad, etcétera.

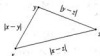


Figura 1.54

Los conceptos geométricos de recta, plano, círculo y circunferencia, esfera en \mathbb{R}^2 o \mathbb{R}^3 tienen sus análogos en \mathbb{R}^n :

Rectas Dados $x, y \in \mathbb{R}^n$, $x \neq y$, la recta que pasa por x e y es el conjunto de puntos de \mathbb{R}^n dados por la expresión

$$(1-t)x + ty, \quad t \in \mathbb{R} \quad (1.115)$$

Poniendo $u = y - x$, (1.115) se puede escribir en la forma

$$x + tu, \quad t \in \mathbb{R} \quad (1.116)$$

Las rectas $r \equiv \{(1-t)x + ty\}$ y $s \equiv \{(1-t)\bar{x} + t\bar{y}\}$ son *paralelas* si los vectores $y - x$ e $\bar{y} - \bar{x}$ son uno múltiplo escalar del otro. Si $x = (0, \dots, 0)$ e $y \neq 0$, la recta $\{ty, t \in \mathbb{R}\}$ pasa por el origen y es el subespacio vectorial de dimensión 1 generado por el vector y . La recta que pasa por x e y es paralela al subespacio $\{t(y - x)\}$.

Las rectas $\{x + t(y - x)\}$ y $\{\bar{x} + t(\bar{y} - \bar{x})\}$ que pasan por el punto x son *ortogonales* si los vectores $y - x$ e $\bar{y} - \bar{x}$ son ortogonales.

El *segmento* (de recta) que une x e y es el conjunto de puntos de \mathbb{R}^n de la forma

$$[x, y] = \{(1-t)x + ty, \quad t \in [0, 1]\} \quad (1.117)$$

o, equivalentemente,

$$[x, y] = \{x + t(y - x), \quad t \in [0, 1]\}$$

Un conjunto $S \subset \mathbb{R}^n$ se dice que es *convexo* si el segmento de recta que une dos cualesquiera de sus puntos está contenido en S .

(Véanse las figuras 1.10, 1.11 y 1.12.)

Hiperplanos Un hiperplano es la generalización de plano en \mathbb{R}^3 . Es, pues, la unión de todas las rectas que pasan por un punto dado p y son ortogonales a

una recta dada que pasa por p , condición que se expresa por la misma ecuación que en \mathbb{R}^3 :

$$\langle \vec{n}, x - p \rangle = 0 \quad (1.118)$$

donde $\vec{n}, p \in \mathbb{R}^n$ están dados. Un hiperplano en \mathbb{R}^n es, por tanto, un conjunto de la forma

$$\{x; \langle \vec{n}, x \rangle = d\} \quad (1.119)$$

con $\vec{n} \in \mathbb{R}^n, \vec{n} \neq 0, d \in \mathbb{R}$ dados ($d = \langle \vec{n}, p \rangle$). \vec{n} y d están determinados salvo un factor escalar, ya que $\{x; \langle \vec{n}, x \rangle = d\} \equiv \{x; \langle c\vec{n}, x \rangle = cd\}$. Desplegando el producto escalar de (1.119) vemos que la ecuación de un hiperplano es de la forma

$$a_1 x_1 + \dots + a_n x_n = d \quad (1.120)$$

con $a_1, \dots, a_n, d \in \mathbb{R}$ dados.

Para $n = 1$, un hiperplano es un punto; para $n = 2$, una recta, y para $n = 3$, un plano.

Dos hiperplanos de la forma $\{x; \langle \vec{n}, x \rangle = d\}$ y $\{x; \langle \vec{n}, x \rangle = \bar{d}\}$, con $\bar{d} \neq d$, son paralelos (visualícese en \mathbb{R}^3). Si $\bar{d} = 0$, $\{x; \langle \vec{n}, x \rangle = 0\}$ es un subespacio vectorial de \mathbb{R}^n de dimensión $n - 1$ (compruébese como ejercicio). Así pues, todo hiperplano de \mathbb{R}^n es el "trasladado paralelamente" de un subespacio vectorial de dimensión $n - 1$.

Semiespacios Los conjuntos $\{x; \langle \vec{n}, x \rangle \geq d\}$ y $\{x; \langle \vec{n}, x \rangle \leq d\}$ son los *semiespacios cerrados*, positivo y negativo, respectivamente, determinados por el hiperplano $H = \{x; \langle \vec{n}, x \rangle = d\}$; su unión es todo \mathbb{R}^n . Los correspondientes *semiespacios abiertos* son $\{x; \langle \vec{n}, x \rangle > d\}$ y $\{x; \langle \vec{n}, x \rangle < d\}$; su unión es $\mathbb{R}^n \setminus H$. (Fig. 1.55).

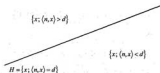


Figura 1.55

Esferas y bolas en \mathbb{R}^n Dados $x_0 \in \mathbb{R}^n$ y $\varepsilon > 0$, el conjunto

$$\{x \in \mathbb{R}^n; |x - x_0| = \varepsilon\} \quad (1.121)$$

se dice que es la *esfera* $(n-1)$ -dimensional de centro x_0 y radio ε . Para $n=1$, es el par de puntos $x_0 \pm \varepsilon$; para $n=2$, una circunferencia y para $n=3$, una esfera (superficie esférica).

El conjunto

$$B_\varepsilon(x_0) = \{x \in \mathbb{R}^n; |x - x_0| < \varepsilon\} \quad (1.122)$$

es la *bola abierta* de centro x_0 y radio ε . Si $n=1$ se trata del intervalo abierto $(x_0 - \varepsilon, x_0 + \varepsilon)$, si $n=2$, de un círculo (sin la circunferencia que lo limita) y si $n=3$, de una esfera sólida sin la superficie esférica que lo limita.

Si en (1.122) la desigualdad no es estricta, o sea, la condición se expresa con \leq , se tiene la *bola cerrada* $\bar{B}_\varepsilon(x_0)$. En el capítulo siguiente se justificarán estos calificativos de abierto y cerrado que se han aplicado a semiespacios y bolas.

Representación gráfica de funciones de más de dos variables

Una función lineal de n variables es de la forma

$$y = f(x_1, \dots, x_n) = a_1 x_1 + \dots + a_n x_n + b \quad (1.123)$$

La gráfica de una de estas funciones, o sea, el conjunto

$$\{(x, y) \in \mathbb{R}^{n+1}; x \in \mathbb{R}^n, y = a_1 x_1 + \dots + a_n x_n + b\}$$

es un hiperplano en el espacio $(n+1)$ -dimensional \mathbb{R}^{n+1} . El coeficiente a_j , $j=1, \dots, n$, da la variación que sufre y como consecuencia de un cambio unitario en la variable x_j (manteniéndose constantes las demás variables). La variación que experimenta y cuando nos desplazamos en el espacio \mathbb{R}^n desde un punto $x_0 = (x_1^0, \dots, x_n^0)$ hasta $x_0 + u$, siendo $u = (u_1, \dots, u_n)$ un vector unitario ($u_1^2 + \dots + u_n^2 = 1$), viene dada por

$$f(x_0 + u) - f(x_0) = a_1 u_1 + \dots + a_n u_n = \langle (a_1, \dots, a_n), (u_1, \dots, u_n) \rangle \quad (1.124)$$

La variación máxima se tendrá cuando

$$u = \frac{(a_1, \dots, a_n)}{\sqrt{a_1^2 + \dots + a_n^2}}$$

o sea, en la dirección del vector (a_1, \dots, a_n) , el cual recibe el nombre de *gradiente* de la función $y = a_1 x_1 + \dots + a_n x_n + b$. El conjunto de nivel k de esta función es el conjunto

$$\{x \in \mathbb{R}^n; a_1 x_1 + \dots + a_n x_n + b = k\} \quad (1.125)$$

Así pues, los conjuntos de nivel de una función lineal de n variables son hiperplanos en \mathbb{R}^n . Obsérvese que si $n = 3$ se trata de *planos* en \mathbb{R}^3 y por eso para las funciones de tres variables es viable una representación gráfica mediante sus conjuntos de nivel.

De la ecuación de un hiperplano de nivel

$$(\text{grad } f, x) = k - b \quad (1.126)$$

vemos que

$$(\text{grad } f, x^2 - x^1) = 0 \quad (1.127)$$

cualesquiera que sean x^1, x^2 del hiperplano. Por lo tanto, el *gradiente de f* es ortogonal a sus hiperplanos de nivel (Fig. 1.56).

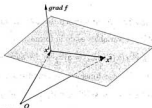


Figura 1.56

La gráfica de una función $y = f(x_1, \dots, x_n)$ no lineal es el conjunto

$$\{(x, y) \in \mathbb{R}^{n+1}; x \in \mathbb{R}^n, y = f(x_1, \dots, x_n)\}$$

Bajo condiciones muy generales sobre la función f , dicha gráfica es una *hipersuperficie* o *variedad de dimensión n* en el espacio \mathbb{R}^{n+1} .

Como para las funciones lineales, el conjunto de nivel k de una función $y = f(x_1, \dots, x_n)$ es

$$\{(x_1, \dots, x_n) \in \mathbb{R}^n; f(x_1, \dots, x_n) = k\} \quad (1.128)$$

Para $n = 3$ todavía cabe una visualización de estos conjuntos. Por ejemplo, los conjuntos de nivel de la función

$$w = x^2 + y^2 + z^2$$

son las esferas de centro el origen de \mathbb{R}^3 , además del propio origen. Pero para $n > 3$ no es posible, naturalmente, tal visualización. Bajo condiciones muy generales sobre la función f , los conjuntos de nivel responden a la definición abstracta de *hipersuperficie* o *variedad de dimensión $n - 1$* en \mathbb{R}^n .

1.5 Problemas

1. Indicar el dominio y el recorrido de las siguientes funciones

a) $f(x, y) = e^{x/y}$

b) $f(x, y) = \sqrt{4 - x^2 - 4y^2}$

c) $f(x, y) = \frac{\ln(x^2 - y)}{y}$

d) $f(x, y) = \sqrt{\frac{1 - x - 2y}{x}}$

e) $f(x, y) = \arcsen(x^2 + y^2)$

f) $f(x, y) = \ln(xy - 2)$

g) $f(x, y) = \ln x + \ln y$

h) $f(x, y) = x^y$

i) $f(x, y) = \sqrt{x^2 + 3y^2 - 6}$

j) $f(x, y) = \sqrt{\frac{1 - (x-1)^2 - y^2}{y - x - 1}}$

k) $f(x, y, z) = x^2 + y^2 + z^2$

l) $f(x, y, z) = \sqrt[4]{1 - x^2 - y^2 - z^2}$

2. Dibujar las curvas de nivel de las siguientes funciones

$f(x, y) = xy$

$f(x, y) = e^{xy}$

$f(x, y) = x + y$

$f(x, y) = |x| + |y|$

$f(x, y) = \max(|x|, |y|)$

$f(x, y) = x^4 + y^2$

$f(x, y) = (x^2 + y^2)^2$

$f(x, y) = x^2$

$f(x, y) = y^2$

3. Para las siguientes funciones, dibujar las curvas correspondientes a los niveles 0 y 1, y las curvas de nivel que pasan por (0,0) y por (1,1)

$f(x, y) = x^2$

$f(x, y) = x - y^2$

$f(x, y) = x^2 + y^2$

$f(x, y) = (x^2 + y^2)^2$

$f(x, y) = xy$

$f(x, y) = y - \sin x$

4. (a) Obtener los puntos de corte con los ejes y dibujar la gráfica de la función
- $f(x, y) = 2 + x + 2y$
- .

- (b) Dibujar las superficies de nivel de
- $f(x, y, z) = x + y - z$
- .

5. Resolver gráficamente los siguientes problemas de optimización:

a)
$$\begin{cases} \text{opt } \{2x_1 + 2x_2\} \\ \text{s.a.} \\ x_1 \leq 3 \\ 3x_1 - 2x_2 \geq -6 \\ 3x_1 + x_2 \geq 3 \end{cases}$$

b)
$$\begin{cases} \text{opt } \{12x_1 + 4x_2\} \\ \text{s.a.} \\ x_1 \geq \frac{1}{2}, x_2 \leq 4 \\ x_1 + x_2 \geq 2 \\ x_1 - x_2 \leq 0 \end{cases}$$

$$c) \begin{cases} \text{opt } \{-4x_1 + 5x_2 + 6\} \\ \text{s.a.} \\ x_1 \geq 2x_2 \\ x_1 + x_2 \geq 1 \\ x_1 - x_2 \leq 3 \end{cases}$$

$$d) \begin{cases} \text{opt } \{x_1 + x_2\} \\ \text{s.a.} \\ x_1^2 + 2x_2^2 \leq 2 \end{cases}$$

$$e) \begin{cases} \text{opt } \{-x_1^2 - x_2^2\} \\ \text{s.a.} \\ 1 \leq x_1 + x_2 \end{cases}$$

$$f) \begin{cases} \text{opt } \{x_1 \cdot x_2\} \\ \text{s.a.} \\ x_1 \geq 0, x_2 \geq 0 \\ 2x_1 + 3x_2 - 5 \leq 0 \end{cases}$$

$$g) \begin{cases} \text{opt } \{(x_1 + 1)^2 + (x_2 + 1)^2\} \\ \text{s.a.} \\ x_1^2 - x_2 \leq 2 \\ 2x_2 - x_1 \leq 2 \\ x_1 \geq 0 \\ x_2 \geq 0 \end{cases}$$

$$h) \begin{cases} \text{opt } \{x_1 + 3x_2 + 3x_3\} \\ \text{s.a.} \\ x_2 + x_3 \leq 3 \\ x_1 - x_2 \geq 0 \\ x_2 \geq 1 \\ 3x_1 + x_2 \leq 15 \end{cases}$$

Capítulo 2

Continuidad

En el primer apartado se exponen de un modo informal las ideas fundamentales acerca de la continuidad de una función con el objetivo de dar una primera aproximación al concepto y de motivar el desarrollo riguroso de dichas ideas que, a partir de la noción de límite, se expone en el segundo apartado y se profundiza en el capítulo 7 con que comienza la segunda parte del libro.

2.1 Generalidades. Conjuntos abiertos y cerrados

La *continuidad* es el primer nivel de regularidad de una función. Se recordará que la idea intuitiva de continuidad de una función de una variable es la de que a puntos próximos, la función hace corresponder valores próximos, o sea, que la función no varía bruscamente. La idea es la misma para funciones de varias variables y queda reflejada en los ejemplos del apartado 1.2 por el hecho de que los puntos $(x, y, f(x, y))$ y $(x_0, y_0, f(x_0, y_0))$ de la superficie representativa de f están próximos cuando (x, y) y (x_0, y_0) están próximos en el plano (son todas ellas superficies “suaves”, que no presentan variaciones bruscas). Otra forma de expresarlo es diciendo que “cuando (x, y) se acerca a (x_0, y_0) , $f(x, y)$ se acerca a $f(x_0, y_0)$ ”: ésta será, convenientemente formalizada, la definición de continuidad de la función $f(x, y)$ en el punto (x_0, y_0) . Concretamente, se dirá que $f(x, y)$ es *continua* en un punto (x_0, y_0) de su dominio si para cada $\varepsilon > 0$ existe un correspondiente $\delta > 0$ tal que para todo punto $(x, y) \in \text{Dom } f$ que esté a una distancia de (x_0, y_0) menor que δ :

$$\sqrt{(x - x_0)^2 + (y - y_0)^2} < \delta$$

se verifica que

$$|f(x, y) - f(x_0, y_0)| < \varepsilon$$

Se observará que esta definición en términos ε - δ (en la que se sobreentiende que ε se toma pequeño, arbitrariamente pequeño, lo que implicará normalmente que el correspondiente δ sea también cada vez más pequeño) es la misma que para funciones de una variable, con la modificación obvia de que la proximidad de puntos en \mathbb{R}^2 hay que medirla con la distancia en el plano, $\sqrt{(x-x_0)^2 + (y-y_0)^2}$, que sustituye a la distancia en la recta dada por el valor absoluto $|x-x_0|$. Nótese que $|f(x,y) - f(x_0,y_0)|$ sí es un valor absoluto, pues la imagen de f está en \mathbb{R} . (Fig. 2.1). Se dice que una función es continua en un conjunto A si es continua en todo punto de A .

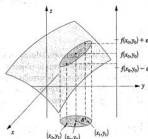


Figura 2.1

Como veremos, la mayoría de las funciones que usualmente aparecen en las aplicaciones son continuas en todos o casi todos los puntos de su dominio de definición. Así, los polinomios como

$$x^2y + 3xy^2 - 2xy + 5$$

son funciones continuas en todo punto (x,y) del plano (un polinomio en dos variables es una suma de términos de la forma cx^ry^s , con c constante y r y s números enteros ≥ 0 ; y análogamente para más de dos variables). Las funciones racionales (cocientes de polinomios) como

$$\frac{5x^3y^2 - 2x^2y^2 + 4x - y}{x^2 - y^2}$$

son continuas en todo punto (x,y) que no anule al denominador (en el ejemplo, excepto en los puntos de las rectas $y = x$ e $y = -x$). También son continuas en todo punto de \mathbb{R}^2 funciones como e^{xy} , $\sin(x^2 + y^2)$, etc., que resultan de componer una función continua de dos variables con una función continua de \mathbb{R} en $\mathbb{R} : (x,y) \rightarrow xy \rightarrow e^{xy}$, etc..

Al estudiar funciones de una variable, normalmente se toma como dominio de una función un intervalo —en particular, todo \mathbb{R} — o acaso la unión de un cierto número de intervalos; en varias variables, sin embargo, los dominios de las funciones pueden ser conjuntos bastante más complicados y por eso se hace necesario un estudio de ciertos aspectos de los conjuntos de \mathbb{R}^n directamente relacionados con el concepto de continuidad; veremos aquí los primeros elementos y completaremos dicho estudio en el apartado que sigue y en el capítulo 7.

El conjunto del plano análogo a un intervalo abierto en \mathbb{R} es el *círculo* (o *disco*) *abierto* —sin la circunferencia que lo limita—

$$B_\varepsilon(x_0, y_0) = \left\{ (x, y) \in \mathbb{R}^2; \sqrt{(x - x_0)^2 + (y - y_0)^2} < \varepsilon \right\}$$

(Fig. 2.2). Llamaremos también a $B_\varepsilon(x_0, y_0)$ ε -*entorno* de (x_0, y_0) . En general, un *entorno* de (x_0, y_0) es un conjunto que contiene un ε -entorno de (x_0, y_0) .



Figura 2.2

El análogo a un intervalo cerrado es el *círculo cerrado* (con la circunferencia)

$$\bar{B}_\varepsilon(x_0, y_0) = \left\{ (x, y) \in \mathbb{R}^2; \sqrt{(x - x_0)^2 + (y - y_0)^2} \leq \varepsilon \right\}$$

Son éstos conjuntos muy especiales, pero con su ayuda se describen conjuntos más generales que aparecen como dominios de funciones.

Un conjunto $S \subset \mathbb{R}^2$ se dice que es *acotado* si se puede incluir en algún círculo centrado en el origen, o sea, si existe $r > 0$ tal que $(x, y) \in B_r(0, 0)$ para todo $(x, y) \in S$ (Fig. 2.3).

Un punto (x_0, y_0) de un conjunto A de \mathbb{R}^2 se dice que es un *punto interior* de A si existe un $\varepsilon > 0$ tal que el círculo $B_\varepsilon(x_0, y_0)$ está totalmente contenido en A . Un punto (x_0, y_0) se dice que es *punto frontera* de A si *toda* círculo $B_\varepsilon(x_0, y_0)$ centrado en (x_0, y_0) contiene puntos de A y puntos del conjunto complementario de A (Fig. 2.4; defínase, a la vista de la figura, *punto exterior* a A).

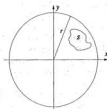


Figura 2.3

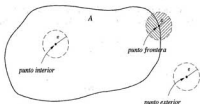


Figura 2.4

El conjunto de puntos interiores de A se denomina *interior* de A y se representa por $\overset{\circ}{A}$ o $\text{int } A$. El conjunto de todos los puntos frontera es la *frontera* de A ; se denota por ∂A o $\text{Fr } A$. Obsérvese que A y su conjunto complementario tienen la misma frontera.

Por definición, un punto interior es necesariamente un punto de A , es decir, $\text{int } A \subset A$, pero un punto frontera puede pertenecer a A o al complementario de A . Si todo punto de A es punto interior, o sea, si $A = \text{int } A$, se dice que A es un *conjunto abierto* (así pues, un conjunto abierto no contiene a ninguno de sus puntos frontera). Si A contiene a todos sus puntos frontera, o sea, si $A \supset \partial A$, se dice que A es un *conjunto cerrado*. El círculo abierto $B_e(x_0, y_0)$ es, como adelantaba su denominación, un conjunto abierto, según se muestra en la figura 2.5 y se demuestra analíticamente más adelante.

La frontera de $B_e(x_0, y_0)$ es la circunferencia $(x - x_0)^2 + (y - y_0)^2 = e^2$ que lo limita; y también lo es del círculo cerrado $\bar{B}_e(x_0, y_0)$, que, en consecuencia, es un conjunto cerrado.

Compruébese que un conjunto es cerrado si y sólo si su complementario

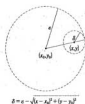


Figura 2.5

es abierto.

Nótese que si $A = \mathbb{R}^2$, el plano entero, todos sus puntos son interiores, con lo que \mathbb{R}^2 es abierto según la definición anterior. Y por tratarse del plano entero, no hay puntos frontera de \mathbb{R}^2 , o sea, $\partial\mathbb{R}^2$ es el conjunto vacío \emptyset ; si, como es usual, se considera al vacío \emptyset como un subconjunto más del espacio total \mathbb{R}^2 , vemos que \mathbb{R}^2 contiene al conjunto de sus puntos frontera, es decir, que \mathbb{R}^2 es cerrado según la definición anterior. Así pues, \mathbb{R}^2 es simultáneamente abierto y cerrado, y como \mathbb{R}^2 y \emptyset son complementarios, lo mismo ocurre para el subconjunto vacío \emptyset . Son los únicos conjuntos de \mathbb{R}^2 para los que eso ocurre. (Véase el apartado 7.7; no es una cuestión trivial).

En la vida ordinaria una ventana o está abierta o está cerrada, pero ¡atención! hay conjuntos que no son abiertos ni cerrados: contienen a algunos de sus puntos frontera pero no a todos, como por ejemplo

$$\{(x, y); x^2 + y^2 < 1\} \cup \{(x, y); x^2 + y^2 = 1, y \geq 0\}$$

(Fig. 2.6)

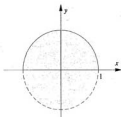


Figura 2.6

En la definición de continuidad, $(x_0, y_0) \in \text{Dom } f$ puede ser un punto interior o un punto frontera. Si (x_0, y_0) es interior, f es continua en (x_0, y_0) si dado $\varepsilon > 0$ existe $\delta > 0$ tal que para todo $(x, y) \in B_\delta(x_0, y_0)$ se tiene $|f(x, y) - f(x_0, y_0)| < \varepsilon$ (pues se puede elegir δ suficientemente pequeño para que $B_\delta(x_0, y_0) \subset \text{Dom } f$). Si (x_0, y_0) es punto frontera, hay que mantener el enunciado original: dado $\varepsilon > 0$ existe $\delta > 0$ tal que para todo $(x, y) \in B_\delta(x_0, y_0) \cap \text{Dom } f$ se tiene $|f(x, y) - f(x_0, y_0)| < \varepsilon$ (pues f sólo se puede evaluar en puntos de $\text{Dom } f$).

2.2 Límite y continuidad de una función

Introduzcamos un conjunto más asociado a A antes de dar el concepto de *límite* de una función de varias variables:

Definición 2.1 La *adherencia* o *cierre* del conjunto A , que se denota por \bar{A} , es la unión de su interior y su frontera: $\bar{A} = \text{int } A \cup \partial A$.

Nótese que, de hecho, $\bar{A} = A \cup \partial A$, y de ahí el nombre: \bar{A} se obtiene “adhiriendo” a A todos sus puntos frontera; compruébese esta afirmación como ejercicio; por otra parte, es claro que A es cerrado si y sólo si $A = \bar{A}$, o sea, coincide con su cierre.

En el caso de $B_\varepsilon(x_0, y_0)$ y $\bar{B}_\varepsilon(x_0, y_0)$ (nótese que $\bar{B}_\varepsilon(x_0, y_0)$ es la adherencia de $B_\varepsilon(x_0, y_0)$ y por eso la notación adoptada desde el principio) la *frontera* corresponde exactamente a la idea intuitiva que su nombre sugiere; en otros casos, ello no es tan intuitivo. Sea, por ejemplo, el conjunto formado por el círculo unidad $B_1(0, 0)$ y el punto $(2, 0)$ (Fig. 2.7).

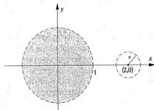


Figura 2.7

El punto $(2, 0)$ es un punto frontera: para cualquier $\varepsilon > 0$, $B_\varepsilon(2, 0)$ contiene al menos un punto de A , el propio $(2, 0)$, y, por supuesto, puntos de

CA. Así pues

$$\partial A = \{(x, y); x^2 + y^2 = 1\} \cup \{(2, 0)\}$$

En general, un punto $(x_0, y_0) \in A$ se dice que es un *punto aislado* si posee un ε -entorno en el que el único punto de A es él mismo, o sea, si existe $\varepsilon > 0$ tal que $B_\varepsilon(x_0, y_0) \cap A = \{(x_0, y_0)\}$. Los puntos aislados de A , caso de existir, son siempre puntos frontera de A , como en el ejemplo anterior.

Un punto de \bar{A} puede o no pertenecer a A pero en cualquiera de los dos casos cumple (sea de $\text{int } A$ o de ∂A) la condición de que en todo ε -entorno suyo existe algún punto de A : en el caso especial de un punto aislado, él mismo cumple trivialmente esa condición. Si $(x_0, y_0) \in \bar{A}$ no es punto aislado de A , entonces *todo* ε -entorno suyo contiene puntos de A —infinitos, de hecho; véase el problema 7 del capítulo 7— distintos del propio (x_0, y_0) ; se dice en este caso que (x_0, y_0) es un *punto de acumulación* de A . Así pues, \bar{A} consta de puntos de acumulación y, eventualmente, puntos aislados.

La idea intuitiva de que una función $f(x, y)$ tiene límite l en el punto (x_0, y_0) , lo que se expresa por

$$\lim_{(x,y) \rightarrow (x_0,y_0)} f(x,y) = l$$

es la de que al aproximarse a (x_0, y_0) a través de puntos (x, y) del dominio de f —distintos del propio (x_0, y_0) —, los valores de $f(x, y)$ se aproximan al valor l . Para concretar esta idea necesitamos que f esté definida en puntos arbitrariamente próximos a (x_0, y_0) , o sea, que haya puntos de $\text{Dom } f$ a cualquier distancia, por pequeña que sea, de (x_0, y_0) . Por otro lado, no queremos excluir la posibilidad de que $(x_0, y_0) \notin \text{Dom } f$, o sea, de que f no esté definida en el propio (x_0, y_0) , como ocurre, por ejemplo, en la definición de derivada de una función de una variable

$$f'(x_0) = \lim_{x \rightarrow x_0} \frac{f(x) - f(x_0)}{x - x_0}$$

en donde el cociente del que se toma límite es una función que no está definida en x_0 . (Y si $(x_0, y_0) \in \text{Dom } f$, obsérvese que el valor de f en el propio (x_0, y_0) , $f(x_0, y_0)$, es irrelevante en el proceso de aproximación descrito, pues hemos impuesto explícitamente la cláusula "... distintos del propio (x_0, y_0) " para los puntos donde se evalúa f .)

Resulta claro, entonces, que para lograr esos objetivos, (x_0, y_0) ha de ser justamente un punto de acumulación del dominio de f . Ponemos entonces la siguiente

Definición 2.2 Sea $f: D \subset \mathbb{R}^2 \rightarrow \mathbb{R}$ y sea (x_0, y_0) un punto de acumulación de D . Se dice que $f(x, y)$ tiene un límite $l \in \mathbb{R}$ cuando (x, y) tiende

a (x_0, y_0) , lo que se expresa por

$$\lim_{(x,y) \rightarrow (x_0,y_0)} f(x,y) = l \quad (2.1)$$

si para todo $\varepsilon > 0$ existe $\delta > 0$ tal que las relaciones

$$(x,y) \in D, 0 < \sqrt{(x-x_0)^2 + (y-y_0)^2} < \delta \quad (2.2)$$

implican $|f(x,y) - l| < \varepsilon$.

En lugar de la expresión (2.1) se escribe también " $f(x,y) \rightarrow l$ cuando $(x,y) \rightarrow (x_0,y_0)$ ". Obsérvese que la condición anteriormente mencionada " (x,y) distinto de (x_0,y_0) " queda muy sucintamente recogida en la primera de las desigualdades de (2.2); es importante insistir en este punto pues la utilidad del concepto de límite reside en recoger la información aportada por los puntos de D próximos a (x_0,y_0) , pero sin contar a éste. (La definición de límite se puede formular de manera equivalente con el signo " \leq " en lugar de " $<$ " en la segunda igualdad de (2.2) y en $|f(x,y) - l| \leq \varepsilon$. Reflexiónese sobre esta afirmación.)

Vemos que la definición de límite es la misma que para funciones de una variable; y no sólo la definición, sino también los resultados básicos sobre límites y continuidad y sus demostraciones son exactamente los mismos que para funciones de una variable. Nos limitaremos, por ello, aquí a enunciarlos e invitamos al lector a transcribir a la situación presente las demostraciones del curso de funciones de una variable.

En primer lugar, el límite, de existir, es único. Además se satisfacen las leyes algebraicas conocidas:

Teorema 2.1 Si $\lim_{(x,y) \rightarrow (x_0,y_0)} f(x,y) = a$, $\lim_{(x,y) \rightarrow (x_0,y_0)} g(x,y) = b$, entonces

$$\lim_{(x,y) \rightarrow (x_0,y_0)} (f(x,y) + g(x,y)) = a + b$$

$$\lim_{(x,y) \rightarrow (x_0,y_0)} f(x,y)g(x,y) = ab$$

$$\lim_{(x,y) \rightarrow (x_0,y_0)} cf(x,y) = ca, \quad c \text{ una constante cualquiera}$$

$$\lim_{(x,y) \rightarrow (x_0,y_0)} \frac{f(x,y)}{g(x,y)} = \frac{a}{b} \quad (\text{si } b \neq 0)$$

A pesar de estas semejanzas, el salto de una a dos (o más) variables supone un cambio en la manera de aproximarse a un punto que tiene importantes consecuencias en la determinación de límites: al tratar de establecer si una función de una variable $f(x)$ tiene límite en un punto $x_0 \in \mathbb{R}$, hay que

analizar qué ocurre al aproximarse a x_0 en la única dirección que cabe en \mathbb{R} (acaso distinguiendo entre uno y otro lado si se trata de límites laterales; si éstos existen y coinciden, entonces existe $\lim_{x \rightarrow x_0} f$). Para una función de dos variables, la aproximación $(x, y) \rightarrow (x_0, y_0)$ se entiende que tiene lugar de cualquier modo según la distancia del plano; desde luego, según cualquier dirección rectilínea, y así se tiene que si existe límite ha de coincidir con el límite al aproximarse a (x_0, y_0) a lo largo de cualquier recta que pase por (x_0, y_0) (Fig. 2.8).

Proposición 2.1 Si existe $\varepsilon > 0$ tal que todos los puntos de $B_\varepsilon(x_0, y_0)$ pertenecen a D excepto quizás el propio (x_0, y_0) , y

$$l = \lim_{(x,y) \rightarrow (x_0,y_0)} f(x,y)$$

entonces, para todo vector unitario $u = (u_1, u_2) \in \mathbb{R}^2$, $u_1^2 + u_2^2 = 1$, se tiene

$$\lim_{t \rightarrow 0} f(x_0 + tu_1, y_0 + tu_2) = l \quad (2.3)$$

Los puntos $(x_0 + tu_1, y_0 + tu_2) = (x_0, y_0) + t(u_1, u_2)$, $t \in \mathbb{R}$, describen la recta que pasa por (x_0, y_0) y tiene a u por vector dirección, y el límite (2.3) es el límite de la función de una variable

$$t \mapsto f(x_0 + tu_1, y_0 + tu_2) \in \mathbb{R}.$$

(generalmente, una dirección en \mathbb{R}^n se da mediante un vector unitario con el objeto de simplificar los cálculos; en \mathbb{R}^2 , es frecuente también dar los vectores unitarios en la forma $(\cos \theta, \sin \theta)$, representando $\theta \in [0, 2\pi)$ el ángulo que forma la recta —orientada— con el semieje positivo de abscisas).

Para probar (2.3) basta tener en cuenta en la definición 2.2 que

$$|(x_0 + tu_1, y_0 + tu_2) - (x_0, y_0)| = \sqrt{t^2 u_1^2 + t^2 u_2^2} = |t|$$

Pero no basta con considerar únicamente las aproximaciones a (x_0, y_0) según líneas rectas, es decir, no es suficiente para que exista

$$\lim_{(x,y) \rightarrow (x_0,y_0)} f(x,y)$$

el que existan y coincidan los límites según cualquier recta que pasa por (x_0, y_0) (o, en otras palabras, no es cierto el recíproco de la proposición 2.1), según muestra el siguiente

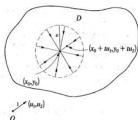


Figura 2.8

EJEMPLO 1. Sea

$$f(x, y) = \frac{xy^2}{x^2 + y^4}, \quad (x, y) \neq (0, 0)$$

y sea $(x_0, y_0) = (0, 0)$.

Tomando cualquier vector unitario (u_1, u_2) se tiene para $t \neq 0$:

$$f(tu_1, tu_2) = \frac{(tu_1)(tu_2)^2}{(tu_1)^2 + (tu_2)^4} = \frac{t^3 u_1 u_2^2}{t^2(u_1^2 + t^2 u_2^4)} = \frac{u_1 u_2^2}{u_1^2 + t^2 u_2^4} t$$

que tiende a 0 cuando $t \rightarrow 0$. (Equivalentemente, las funciones de una variable

$$x \mapsto f(x, mx) = \frac{mx}{1 + m^2 x^2}$$

con $m \in \mathbb{R}$, convergen a 0 cuando $x \rightarrow 0$ cualquiera que sea m , lo mismo que la función $y \mapsto f(0, y) = 0$ cuando $y \rightarrow 0$, restricción esta última de f sobre los semiejes verticales). Así pues, si existiese límite habría de ser $l = 0$. Pero, por otro lado, todo δ -entorno de $(0, 0)$ contiene puntos de la parábola $x = y^2$ distintos de $(0, 0)$, en los cuales la función vale

$$f(y^2, y) = \frac{y^4}{y^4 + y^4} = \frac{1}{2}$$

y por tanto no existe el límite de f cuando $(x, y) \rightarrow (0, 0)$ pues no se cumple la definición 2.2 para valores $\varepsilon < \frac{1}{2}$ (Fig. 2.9).

La proposición 2.1 es útil, en tanto que condición necesaria, para descartar la existencia de límite cuando el límite según dos rectas es distinto:

EJEMPLO 2. Sea

$$f(x, y) = \frac{y^2}{x^2 + y^2}, \quad (x, y) \neq (0, 0)$$

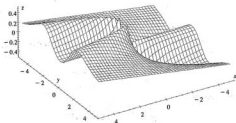


Figura 2.9

Con $(u, v) = (1, 0)$ (o sea, a lo largo del eje x) se tiene para $t \neq 0$

$$f(t, 0) = \frac{0}{t^2 + 0} = 0 \rightarrow 0 \text{ cuando } t \rightarrow 0$$

mientras que con $(u, v) = (0, 1)$ (según el eje y)

$$f(0, t) = \frac{t^2}{0 + t^2} = 1 \rightarrow 1 \text{ cuando } t \rightarrow 0$$

Como esos límites son distintos, no existe el límite de la función f cuando $(x, y) \rightarrow (0, 0)$ (Fig. 2.10).

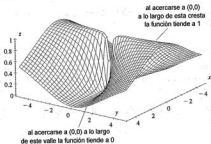


Figura 2.10

También es útil para sugerir un candidato a límite, que es obligado verificar de acuerdo con la definición 2.2:

EJEMPLO 3. Sea

$$f(x, y) = \sqrt{|xy|}$$

Es claro que el límite según cualquier recta que pasa por el origen es 0. Se trata entonces de probar la *conjetura*

$$\lim_{(x,y) \rightarrow (0,0)} \sqrt{|xy|} = 0$$

De la desigualdad

$$(x \pm y)^2 = x^2 \pm 2xy + y^2 \geq 0$$

se tiene

$$|xy| \leq \frac{1}{2}(x^2 + y^2)$$

de donde

$$f(x, y) = |xy|^{\frac{1}{2}} \leq \frac{1}{\sqrt{2}}(x^2 + y^2)^{\frac{1}{2}}$$

Dado $\varepsilon > 0$, basta tomar $\delta = \sqrt{2}\varepsilon$. En este tipo de argumentos es más cómodo a veces utilizar coordenadas polares:

$$\left| \sqrt{|xy|} - 0 \right| = \sqrt{|xy|} = \sqrt{|r \cos \theta \cdot r \sin \theta|} = \sqrt{r^2 \left| \frac{1}{2} \sin 2\theta \right|} \leq \frac{1}{\sqrt{2}} r$$

y se razona del mismo modo teniendo en cuenta que r es precisamente la distancia de (x, y) a $(0, 0)$.

Con mayor razón que en el caso de una variable, sería muy trabajoso recurrir cada vez a la propia definición para el cálculo de un límite. Como allí, lo práctico es disponer de una colección de límites sencillos y utilizar las reglas algebraicas del teorema 2.1. Dos límites realmente simples pero muy útiles son los siguientes:

EJEMPLO 4. Si $f(x, y) = c$ (una constante), entonces

$$\lim_{(x,y) \rightarrow (x_0,y_0)} f(x, y) = c$$

cualquiera que sea $(x_0, y_0) \in \mathbb{R}^2$. En efecto, dado $\varepsilon > 0$, cualquier $\delta > 0$ satisface las condiciones de la definición 2.2, ya que

$$0 < \sqrt{(x - x_0)^2 + (y - y_0)^2} < \delta$$

implica

$$|f(x, y) - f(x_0, y_0)| = |c - c| = 0 < \varepsilon$$

EjemPlo 5.

$$\lim_{(x,y) \rightarrow (x_0,y_0)} x = x_0$$

cualquiera que sea $(x_0, y_0) \in \mathbb{R}^2$. En efecto, dado $\varepsilon > 0$, tomamos $\delta = \varepsilon$; como

$$|x - x_0| \leq \sqrt{(x - x_0)^2 + (y - y_0)^2}$$

se tendrá que

$$0 < \sqrt{(x - x_0)^2 + (y - y_0)^2} < \delta$$

implica

$$|f(x, y) - f(x_0, y_0)| = |x - x_0| < \delta = \varepsilon$$

Análogamente

$$\lim_{(x,y) \rightarrow (x_0,y_0)} y = y_0$$

Con estos dos ejemplos y las reglas que da el teorema 2.1 se tiene que

$$\lim_{(x,y) \rightarrow (x_0,y_0)} P(x, y) = P(x_0, y_0)$$

para todo polinomio $P(x, y)$, cualquiera que sea $(x_0, y_0) \in \mathbb{R}^2$, y que

$$\lim_{(x,y) \rightarrow (x_0,y_0)} \frac{P(x, y)}{Q(x, y)} = \frac{P(x_0, y_0)}{Q(x_0, y_0)}$$

para cualquier función racional y todo $(x_0, y_0) \in \mathbb{R}^2$ tal que $Q(x_0, y_0) \neq 0$.

Comentemos finalmente que otra dificultad adicional que presenta el cálculo explícito de límites de funciones de varias variables es la inexistencia de métodos análogos a la regla de L'Hôpital, herramienta utilísima en el estudio de los límites de funciones de una variable.

La definición de continuidad también es formalmente la misma que para funciones de una variable:

Definición 2.3 $f: D \subseteq \mathbb{R}^2 \rightarrow \mathbb{R}$ se dice que es continua en $(x_0, y_0) \in D$ si

$$\lim_{(x,y) \rightarrow (x_0,y_0)} f(x,y) = f(x_0,y_0) \quad (2.4)$$

En realidad, la condición (2.4) no tiene sentido si (x_0, y_0) no es punto de acumulación de D , o sea, si se trata de un punto aislado de D . Salvamos esta situación excepcional acordando que (2.4) se verifica automáticamente en un punto aislado (x_0, y_0) , o sea, que $f(x, y)$ es por definición continua en todo punto aislado de D (lo que, realmente, no choca con la definición 2.2 si tomamos $l = f(x_0, y_0)$ y, con un $\delta > 0$ suficientemente pequeño, permitimos que la condición de aquella sea, para un punto aislado (x_0, y_0) , " $(x, y) \in D$, $0 \leq \sqrt{(x-x_0)^2 + (y-y_0)^2} < \delta$ implica $|f(x, y) - l| < \varepsilon$ ": es el propio (x_0, y_0) el único punto (x, y) que interviene en la condición, donde se ha cambiado el signo $<$ por \leq , y ésta resulta trivialmente satisfecha). Con esta aclaración, la definición 2.3 se puede expresar en la forma equivalente:

Definición 2.4 $f: D \subseteq \mathbb{R}^2 \rightarrow \mathbb{R}$ se dice que es continua en $(x_0, y_0) \in D$ si para todo $\varepsilon > 0$ existe $\delta > 0$ tal que las relaciones

$$(x, y) \in D, \sqrt{(x-x_0)^2 + (y-y_0)^2} < \delta$$

implican

$$|f(x, y) - f(x_0, y_0)| < \varepsilon$$

Se dice que f es continua en un conjunto $A \subseteq D$ si es continua en cada uno de los puntos de A . Notemos que si f es continua en A , el número $\delta > 0$ correspondiente, en la definición de continuidad, a un $\varepsilon > 0$ dado, dependerá, en general, del punto $(x_0, y_0) \in A$ considerado. Cuando para todo $\varepsilon > 0$ existe un $\delta = \delta(\varepsilon) > 0$ con el que se satisface la definición de continuidad para todo $(x_0, y_0) \in A$, se dice que f es **uniformemente continua** en A . En el capítulo 7 demostraremos que si f es continua en un conjunto cerrado y acotado A , entonces f es uniformemente continua en A (resultado que generaliza el correspondiente a funciones de una variable continuas en un intervalo cerrado y acotado de la recta).

En principio, para establecer que una función es continua hay remitirse, como en el caso de los límites, a la propia definición. Por ejemplo, hemos comprobado en el ejemplo 3 que la función $f(x, y) = \sqrt{|xy|}$ es continua en $(0, 0)$ dado que

$$\lim_{(x,y) \rightarrow (0,0)} \sqrt{|xy|} = 0$$

Veamos que, de hecho, es continua en todo $(x_0, y_0) \in \mathbb{R}^2$. Utilizando coordenadas polares con el polo en (x_0, y_0) y eje polar la recta $y = y_0$:

$$x = x_0 + r \cos \theta, \quad y = y_0 + r \sin \theta$$

se tiene

$$\begin{aligned} \left| \sqrt{|xy|} - \sqrt{|x_0 y_0|} \right| &= \left| \sqrt{|(x_0 + r \cos \theta)(y_0 + r \sin \theta)|} - \sqrt{|x_0 y_0|} \right| \leq \\ &\leq \sqrt{|x_0 y_0| + r|x_0||\sin \theta| + r|y_0||\cos \theta| + \frac{1}{2}r^2|\sin 2\theta|} - \sqrt{|x_0 y_0|} \leq \\ &\leq (\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}) \leq \sqrt{|x_0 y_0|} + \sqrt{r(|x_0| + |y_0|)} + \frac{1}{\sqrt{2}}r - \sqrt{|x_0 y_0|} = \\ &= (|x_0| + |y_0|)^{1/2} r^{1/2} + \frac{1}{\sqrt{2}}r \end{aligned}$$

que tiende a 0 cuando $r \rightarrow 0$ (es decir, para todo $\varepsilon > 0$ existe $\delta > 0$ tal que si $r < \delta$, entonces esa cantidad es menor que ε).

Si bien este tipo de demostración es en ocasiones ineludible, en la mayoría de las aplicaciones la continuidad de una función se establece, como decíamos antes para el cálculo de límites, sobre la base de algunos ejemplos simples y los resultados generales contenidos en los dos teoremas siguientes. El primero de ellos se deriva inmediatamente del teorema 2.1:

Teorema 2.2 Si f y g son continuas en (x_0, y_0) , también lo son $f+g$, fg y cf (c una constante cualquiera). Si, además, $g(x_0, y_0) \neq 0$, entonces f/g es continua en (x_0, y_0) .

Este teorema junto con los ejemplos 4 y 5 implica, como adelantábamos en el apartado 1, que los polinomios $P(x, y)$ son funciones continuas en \mathbb{R}^2 y que una función racional es continua en todo punto que no anule al denominador.

El segundo resultado fundamental es el relativo a la composición de funciones continuas:

Teorema 2.3 Sea $f(x, y)$ continua en (x_0, y_0) .

1. Si g es continua en $f(x_0, y_0)$, entonces $g \circ f$ es continua en (x_0, y_0) (aquí, g es una función de una variable y $g \circ f$ es la función compuesta $(g \circ f)(x, y) = g(f(x, y))$, definida para todo $(x, y) \in \text{Dom } f$ tal que $f(x, y)$ esté en $\text{Dom } g \subseteq \mathbb{R}$).
2. Si $u(t)$ y $v(t)$ son dos funciones definidas en un intervalo $I \subset \mathbb{R}$ tales que $(u(t), v(t)) \in \text{Dom } f$ para todo $t \in I$, son continuas en el punto $t_0 \in I$ y $(u(t_0), v(t_0)) = (x_0, y_0)$, entonces la función de una variable $I \ni t \mapsto f(u(t), v(t))$ es continua en t_0 .

El teorema 2.3 se demuestra siguiendo las pautas del resultado análogo para funciones de una variable.

Con los teoremas 2.2 y 2.3 y las numerosas funciones continuas de una variable que se conocen se establece la continuidad de funciones como e^{xy} , $\sin(x^2 + y^2)$, ..., y muchas otras que aparecen en las aplicaciones. Por ejemplo, la función $f(x, y) = \sqrt{|xy|}$ considerada anteriormente, es el resultado de la composición de funciones

$$(x, y) \mapsto xy \mapsto |xy| \mapsto \sqrt{|xy|}$$

La primera de ellas es una función continua de \mathbb{R}^2 en \mathbb{R} por el teorema 2.3 y las otras dos (la función valor absoluto y la función raíz cuadrada) son funciones continuas de \mathbb{R} en \mathbb{R} conocidas.

Funciones de n variables

La transcripción de estas ideas y resultados a funciones de más de dos variables es mera rutina que, con todo, conviene explicitar por escrito como ejercicio: el papel de los círculos abiertos y cerrados lo juegan en \mathbb{R}^n las bolas abiertas y cerradas introducidas en el apartado 1.4:

$$B_\varepsilon(x_0) = \{x \in \mathbb{R}^n; |x - x_0| < \varepsilon\}$$

$$\bar{B}_\varepsilon(x_0) = \{x \in \mathbb{R}^n; |x - x_0| \leq \varepsilon\}$$

Mediante ellas, se introducen como en \mathbb{R}^2 los conceptos de conjunto acotado, punto interior, punto frontera, conjunto abierto y cerrado, límite de funciones, continuidad, Insistimos en la conveniencia de practicar estos "ejercicios de abstracción". Háganse aquí en especial para las definiciones de límite y continuidad de una función de n variables en un punto.

Demostremos el resultado pendiente de que $B_\varepsilon(x_0)$ es un conjunto abierto: dado $x \in B_\varepsilon(x_0)$, sea $\delta = \varepsilon - |x - x_0|$ (Fig. 2.5). Entonces, si $z \in B_\delta(x)$, se tiene $z - x_0 = (z - x) + (x - x_0)$, y, por la desigualdad triangular,

$$|z - x_0| = |z - x| + |x - x_0| < \delta + |x - x_0| = \varepsilon$$

es decir, $B_\delta(x) \subset B_\varepsilon(x_0)$.

Compruébese que $\partial B_\varepsilon(x_0) = \partial \bar{B}_\varepsilon(x_0) = \{x \in \mathbb{R}^n; |x - x_0| = \varepsilon\}$, la esfera $(n-1)$ -dimensional de centro x_0 y radio ε (véase problema 4 del capítulo 7).

La generalización del teorema 2.3 es:

Teorema 2.4 Sea $f(x)$ continua en $x_0 = (x_1^0, \dots, x_n^0)$.

1. Si g es continua en $f(x_0)$, entonces $g \circ f$ es continua en x_0 (aquí, g es una función de una variable y $g \circ f$ es la función compuesta $(g \circ f)(x) = g(f(x))$, definida para todo $x \in \text{Dom } f \subset \mathbb{R}^n$ tal que $f(x)$ esté en $\text{Dom } g \subseteq \mathbb{R}$).
2. Si $u_1(t_1, \dots, t_m), \dots, u_n(t_1, \dots, t_m)$ son funciones definidas en $D' \subset \mathbb{R}^m$ tales que $(u_1(t), \dots, u_n(t)) \in \text{Dom } f$ para todo $t = (t_1, \dots, t_m) \in D'$, son continuas en $t_0 = (t_1^0, \dots, t_m^0) \in D'$ y $(u_1(t_0), \dots, u_n(t_0)) = x_0$, entonces la función de m variables

$$D' \ni (t_1, \dots, t_m) = t \mapsto f(u_1(t), \dots, u_n(t))$$

es continua en t_0 .

(Inténtese la demostración como ejercicio y consúltese el capítulo 7.)

2.3. Problemas

1. Demostrar, aplicando la definición de límite, las siguientes afirmaciones:

a) $\lim_{(x,y) \rightarrow (0,0)} x^2 = 0$

b) $\lim_{(x,y) \rightarrow (0,0)} (x^2 + y^2) = 0$

c) $\lim_{(x,y) \rightarrow (0,0)} (x + y^2) = 0$

d) $\lim_{(x,y) \rightarrow (0,0)} xy = 0$

2. Determinar los siguientes límites

a) $\lim_{(x,y) \rightarrow (0,0)} \frac{\sin(x^2 + y^2)}{x^2 + y^2}$

b) $\lim_{(x,y) \rightarrow (0,0)} \frac{x^2}{\sqrt{x^2 + y^2}}$

c) $\lim_{(x,y) \rightarrow (0,0)} \frac{x^2 y}{x^2 + y^2}$

d) $\lim_{(x,y) \rightarrow (0,0)} \frac{(xy)^2}{xy^3 + (x - y)^2}$

e) $\lim_{(x,y) \rightarrow (0,0)} \frac{x^3 + y^3}{x^2 + y^2}$

f) $\lim_{(x,y) \rightarrow (0,0)} \frac{xy}{x^2 + y^2}$

3. (a) Supongamos que existe un límite común l de $f(x, y)$ cuando nos acercamos al origen a lo largo de cualquier dirección $(\cos \theta, \sin \theta)$, $\theta \in [0, 2\pi)$; se tiene, por tanto, que dados $\theta \in [0, 2\pi)$ y $\varepsilon > 0$, existe $\delta = \delta(\varepsilon, \theta) > 0$ tal que si $\text{dist}((t \cos \theta, t \sin \theta), (0, 0)) < \delta$, entonces

$$|f(t \cos \theta, t \sin \theta) - l| < \varepsilon.$$

Probar que si $\inf_{\theta} \delta(\varepsilon, \theta) = \delta_0(\varepsilon) > 0$ para cada $\varepsilon > 0$, entonces

$$\lim_{(x,y) \rightarrow (0,0)} f(x,y) = 1$$

- (b) Estudiar el límite cuando $(x,y) \rightarrow (0,0)$ de la función

$$f(x,y) = \begin{cases} \frac{y}{x} \sqrt{x^2 + y^2} & \text{si } x \neq 0 \\ 0 & \text{si } x = 0 \end{cases}$$

4. Dada la función

$$f(x,y) = \begin{cases} 1 & \text{si } x^2 < y < 2x^2 \\ 0 & \text{en los demás puntos de } \mathbb{R}^2 \end{cases}$$

se pide:

- Dibujar los conjuntos de nivel.
- Probar que el límite cuando nos aproximamos al origen a lo largo de cualquier recta vale cero (o sea, $\lim_{t \rightarrow 0} f(t \cos \theta, t \sin \theta) = 0$ para todo θ).
- Probar que f tiende a 1 a lo largo de ciertas curvas que pasan por el origen y que, por consiguiente, f no tiene límite en el origen.

5. Estudiar la continuidad de las funciones

$$(a) f(x,y) = \begin{cases} \sqrt{1+xy} & \text{si } xy \geq 0 \\ e^{x+y} & \text{si } xy < 0 \end{cases}$$

$$(b) f(x,y) = \begin{cases} \frac{x^2 y^3}{x^2 + y^2} & \text{si } (x,y) \neq (0,0) \\ a & \text{si } (x,y) = (0,0) \end{cases}$$

$$(c) f(x,y) = \begin{cases} \ln(1+x^2-y^2) & \text{si } x^2 \geq y^2 \\ x+y & \text{si } x^2 < y^2 \end{cases}$$

$$(d) f(x,y) = \begin{cases} \frac{x^2 y}{x^2 + y^2} & \text{si } (x,y) \neq (0,0) \\ 0 & \text{si } (x,y) = (0,0) \end{cases}$$

$$(e) f(x,y) = \begin{cases} \frac{xy}{x^2 + y^2} & \text{si } (x,y) \neq (0,0) \\ 0 & \text{si } (x,y) = (0,0) \end{cases}$$

$$(f) f(x,y) = \begin{cases} \frac{x}{x^2 - y} & \text{si } x^2 - y \neq 0 \\ 0 & \text{si } x^2 - y = 0 \end{cases}$$

6. Sea $f(x, y)$ continua en (x_0, y_0) . Probar que si $f(x_0, y_0) \neq 0$, entonces $f(x, y) \neq 0$ para todo punto (x, y) de un cierto ε -entorno de (x_0, y_0) .

7. Probar que son abiertos los conjuntos

(a) $\{(x, y) \in \mathbb{R}^2; |x| > 1\}$

(b) $\{(x, y) \in \mathbb{R}^2; \sqrt{x^2 + y^2} > 0\}$

(c) $\{(x, y) \in \mathbb{R}^2; |x| > 0\}$

8. ¿Es abierto en \mathbb{R}^2 el eje x ? ¿Es cerrado?

9. Probar que si A y B son dos conjuntos abiertos (resp. cerrados) entonces $A \cup B$ y $A \cap B$ son también conjuntos abiertos (resp. cerrados).

10. Sea $f: \mathbb{R}^2 \rightarrow \mathbb{R}$ una función continua en todo punto de \mathbb{R}^2 . Dados $a, b \in \mathbb{R}$ cualesquiera, demostrar las siguientes afirmaciones:

(a) El conjunto $\{(x, y) \in \mathbb{R}^2; f(x, y) < b\}$ es abierto.

(b) El conjunto $\{(x, y) \in \mathbb{R}^2; f(x, y) > a\}$ es abierto.

(c) El conjunto $\{(x, y) \in \mathbb{R}^2; a < f(x, y) < b\}$ es abierto.

(d) El conjunto $\{(x, y) \in \mathbb{R}^2; f(x, y) \leq b\}$ es cerrado.

(e) El conjunto $\{(x, y) \in \mathbb{R}^2; f(x, y) \geq a\}$ es cerrado.

(f) El conjunto $\{(x, y) \in \mathbb{R}^2; a \leq f(x, y) \leq b\}$ es cerrado.

(g) El conjunto $\{(x, y) \in \mathbb{R}^2; f(x, y) = a\}$ es cerrado.

(h) El conjunto $\{(x, y) \in \mathbb{R}^2; f(x, y) \neq a\}$ es abierto.

(En el capítulo 7 profundizaremos en las cuestiones planteadas en los problemas 9 y 10.)

11. Utilizar los dos problemas anteriores para estudiar si los siguientes conjuntos son abiertos o cerrados:

(a) los semiplanos $\{(x, y) \in \mathbb{R}^2; ax + by < c\}, \{(x, y) \in \mathbb{R}^2; ax + by > c\},$
 $\{(x, y) \in \mathbb{R}^2; ax + by \leq c\}, \{(x, y) \in \mathbb{R}^2; ax + by \geq c\}.$

(b) la recta $\{(x, y) \in \mathbb{R}^2; ax + by = c\}.$

(c) $\{(x, y) \in \mathbb{R}^2; x^2 + y^2 < 1, x > 0\}.$

(d) $\{(x, y) \in \mathbb{R}^2; x^2 + y^2 < 1, x < y < 2x, x > 0\}$

Capítulo 3

Derivadas parciales. Diferenciabilidad

En este capítulo se aborda el tema central del cálculo diferencial: la noción de *derivada* y *diferencial*, y sus propiedades esenciales. Empezamos en el primer apartado repasando someramente el concepto de derivada de una función de una variable, para motivar su extensión al ámbito de varias variables. La interpretación geométrica tradicional de la derivada de una función de una variable en relación con la recta tangente a su curva característica se puede extender sin mayor dificultad para una función de dos variables mediante la consideración del *plano tangente* a la *superficie* que la representa. Esta idea geométrica se traduce analíticamente en la extensión de la idea de *diferencial* a las funciones de dos y más variables. Se pone también el énfasis en los aspectos *computacionales* de estas nociones, tan importantes en las aplicaciones. En los apartados 2 y 3 se exponen con rigor los conceptos y resultados básicos para funciones de dos variables apoyándonos en las consideraciones intuitivas que permiten las representaciones gráficas en el plano y el espacio; se deja para el final de cada uno de ellos la formulación de los correspondientes conceptos y resultados para las funciones de n variables así como la demostración de algún resultado más difícil. El capítulo termina con una serie de notas relativas a ciertos aspectos complementarios o del desarrollo histórico de los conceptos expuestos.

3.1 Introducción

Empezaremos recordando brevemente las ideas básicas que motivan la introducción del cálculo diferencial para funciones de una variable. Uno de los aspectos fundamentales de la descripción de las propiedades de una relación

funcional $y = f(x)$ es estudiar cómo “responde” la variable dependiente y ante variaciones de la variable independiente x . Si, partiendo de un valor dado $x = x_0$, esta variable experimenta una variación o *incremento* h , la variable dependiente y pasará de su valor inicial $y_0 = f(x_0)$ a un nuevo valor $f(x_0 + h)$. La variación o incremento experimentado por y en respuesta al incremento h de x será, pues

$$\text{Incremento de } f \stackrel{\text{def}}{=} f(x_0 + h) - f(x_0)$$

y se denota $\Delta f(x_0, h)$ o simplemente Δf si no se precisa especificar los datos x_0 y h . También se escribe Δy si se quiere poner énfasis en la variable y , en cuyo caso se escribe Δx en vez de h por la misma razón.

Si la relación funcional es lineal, $y = f(x) = ax + b$, es obvio que $f(x_0 + h) - f(x_0) = ah$, o sea, $\Delta y = a\Delta x$. Esto significa que la *variación relativa* o *cociente incremental* $\Delta y/\Delta x$ es independiente tanto de x_0 como de h . Por tanto, tiene sentido decir que, en respuesta a una variación *cualquiera* Δx , a partir de *cualquier* valor inicial x_0 , la variable y se modifica en $a\Delta x$. Desde el punto de vista geométrico, la variación relativa se corresponde con la *pendiente* de la recta, que es un valor perfectamente definido, e independiente del par de puntos de la recta que se utilicen para medirlo.

En el caso de una relación funcional no lineal $y = f(x)$, es precisamente el carácter curvo de su gráfica lo que genera la falta de uniformidad de su pendiente a lo largo de la misma, pues según el par de puntos de la curva que elijamos obtendremos un segmento (llamado *cuerda*) distinto, cuya pendiente, por tanto, también será distinta (si no fuera así, la curva sería una recta). Por esta razón, no se puede hablar de pendiente de una curva si no se especifica qué método se ha seguido para medirla (véase la figura 3.1). En términos analíticos, esto significa que la variación absoluta Δy y la relativa $\Delta y/\Delta x$ dependen, en principio, tanto de x_0 como de Δx .

La observación fundamental que caracteriza a todos los argumentos del cálculo diferencial es la siguiente: Fijemos x_0 de momento. Si bien es cierto que $\Delta y/\Delta x$ depende de Δx , no lo es menos que dicha dependencia va perdiendo importancia a medida que Δx disminuye, lo que se expresa diciendo que $\Delta y/\Delta x$ *tiende a un valor límite cuando Δx tiende a 0*. En términos geométricos, las pendientes $\Delta y/\Delta x$ de las cuerdas determinadas por los puntos (x_0, y_0) y $(x_0 + \Delta x, y_0 + \Delta y)$ tienden a la pendiente de la recta límite de las rectas (llamadas *secantes*) que contienen a las cuerdas. Dicha recta límite es precisamente la *tangente* a la curva $y = f(x)$ en el punto (x_0, y_0) , y por eso se dice que la *tangente es la posición límite de las secantes cuando se mantiene fijo un punto de las cuerdas y el otro se aproxima al primero*.

Resaltemos las dos observaciones que llevamos hechas:

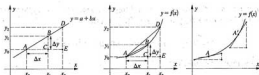


Figura 3.1

- Como, en general, la recta tangente no es secante, la pendiente de la tangente no corresponde en realidad a ninguna variación relativa verdadera ($\Delta y / \Delta x$) de y en respuesta a la variación Δx de x .
- A pesar de esto, la pendiente de la tangente es una buena aproximación a las verdaderas pendientes $\Delta y / \Delta x$, y esta aproximación es tanto mejor cuanto menor sea el incremento Δx , es decir, cuanto más pequeña sea la cuerda considerada.

Estas dos observaciones se pueden resumir de la siguiente forma: *Una curva no es una recta, pero cuanto más de cerca la observemos, más lo parece.*

Llegamos así al concepto de *derivada* introducido en el curso de funciones de una variable:

$$f'(x_0) = \lim_{h \rightarrow 0} \frac{\Delta y}{\Delta x} = \lim_{h \rightarrow 0} \frac{f(x_0 + h) - f(x_0)}{h} \quad (3.1)$$

y su interpretación como pendiente de la curva $y = f(x)$ en el punto (x_0, y_0) , definida ésta como la pendiente de su recta *tangente* en dicho punto. Para interpretar vectorialmente esta definición, llamemos \vec{u}_h al vector de incrementos $(\Delta x, \Delta y) = (h, f(x_0 + h) - f(x_0))$, o sea, la cuerda que une (x_0, y_0) con $(x_0 + \Delta x, y_0 + \Delta y) = (x_0 + h, f(x_0 + h))$ considerada como segmento orientado. Para estudiar el comportamiento de dichas cuerdas cuando $h \rightarrow 0$, primero “ampliamos” o normalizamos los vectores \vec{u}_h de forma que sus componentes horizontales valgan 1, lo que equivale a definir

$$\vec{v}_h \stackrel{\text{def}}{=} \frac{1}{h} \vec{u}_h = \left(1, \frac{f(x_0 + h) - f(x_0)}{h} \right)$$

Vemos así que, supuesta la existencia de derivada, \vec{v}_h tiende a $(1, f'(x_0))$, vector que llamaremos *tangente* a la curva $y = f(x)$ en (x_0, y_0) , si bien la denominación “vector tangente” suele aplicarse, con cierta ambigüedad, a

cualquier vector director de la recta tangente. Deducimos así que la *ecuación vectorial de la recta tangente a $y = f(x)$ en (x_0, y_0)* es

$$(x, y) = (x_0, y_0) + t\vec{u}_0 = (x_0, y_0) + t(1, f'(x_0)),$$

las ecuaciones paramétricas

$$\{x = x_0 + t, \quad y = y_0 + t f'(x_0)\}$$

y la ecuación cartesiana

$$\frac{y - y_0}{x - x_0} = f'(x_0), \quad \text{o bien} \quad y = y_0 + f'(x_0)(x - x_0)$$

La última forma se puede interpretar de la siguiente manera: la función *lineal*

$$l(x) \stackrel{\text{def}}{=} y_0 + f'(x_0)(x - x_0) \quad (3.2)$$

es, de todas las funciones lineales (que son las más simples de las funciones, sin duda), la que mejor aproxima a la función $y = f(x)$ cerca del punto (x_0, y_0) en el sentido de que sus gráficas son tangentes en ese punto. Más adelante perfilaremos esta idea de aproximación mediante la noción de "diferencial".

A veces la variable independiente x es el tiempo, en cuyo caso la relación funcional $y = f(x)$ representa la *evolución temporal* o la *dinámica* de la variable y a lo largo del tiempo. En tal caso, los cocientes incrementales $\Delta y / \Delta x$ representan *velocidades medias*, entendiendo la palabra "velocidad" no en el sentido estricto de "espacio recorrido por unidad de tiempo" sino en un sentido amplio de "variación de y por unidad de tiempo" que mide la *rapidez* con que y varía respecto a x . Cuando el intervalo de tiempo Δx se va reduciendo, la velocidad media se va aproximando a un valor que parece lógico llamar *velocidad instantánea*. Esto nos lleva a la conocida interpretación *dinámica* de la derivada como velocidad (instantánea) del "proceso dinámico" $y = f(x)$.

La derivada y el cálculo incremental aproximado. Diferencial de una función de una variable

Las interpretaciones geométricas y dinámicas de la derivada son muy importantes para entender el concepto y sus consecuencias pero no reflejan con nitidez sus aspectos *computacionales*, los cuales son, sin embargo, la verdadera esencia del cálculo diferencial, que se llama así porque sirve para *calcular* valores, sobre todo *incrementos*. Para entender mejor esta faceta calculística, tan frecuentemente olvidada, describamos brevemente un ejemplo sencillo:

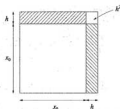


Figura 3.2

Consideremos un cuadrado de lado x y área $f(x) = x^2$, tomemos un valor inicial x_0 del lado y supongamos que éste recibe un incremento h , transformándose así en $x_0 + h$ (figura 3.2).

El área del nuevo cuadrado será

$$f(x_0 + h) = (x_0 + h)^2 = x_0^2 + 2x_0h + h^2$$

El incremento del área asociado al incremento h del lado es

$$\Delta f = f(x_0 + h) - f(x_0) = \{x_0^2 + 2x_0h + h^2\} - x_0^2 = 2x_0h + h^2 \quad (3.3)$$

Esta fórmula, deducida con las reglas elementales del álgebra, es válida para todos los valores de x_0 y de h . La observación que, como comentábamos antes, constituye la idea motriz del cálculo diferencial, es la siguiente: los dos términos que aparecen en la fórmula de Δf no tienen la misma importancia y, cuando nos restringimos a valores pequeños de h , dicha fórmula corresponde a una descomposición de Δf en diferentes escalas de magnitud. Por ejemplo, si $x_0 = 1$ y el incremento h es muy pequeño, el término (de primer orden) $2x_0h = 2h$ es mucho mayor que h^2 (término de orden superior). Así, si $h = 0,1$, vemos que $\Delta f = 2 \times 0,1 + (0,1)^2 = 0,21$ y el término $h^2 = 0,01$ sólo representa el 5% del valor total de Δf . Si $h = 0,001$, la desproporción sería aún más mayor: $h^2 = 10^{-6}$ es sólo el 0,05% de $\Delta f = 0,002001$. La interpretación geométrica de esta desproporción es inmediata, razón por la cual hemos elegido este ejemplo: h^2 representa el área del cuadrado pequeño del ángulo superior derecho en la figura 3.2 mientras que $2x_0h$ es la suma de las áreas de los rectángulos rayados.

¿En qué condiciones tendremos una descomposición análoga de Δf para una función general $f(x)$? Supongamos que $f(x)$ es derivable en x_0 ; si llamamos

$$\delta(h) = f(x_0 + h) - f(x_0) - f'(x_0)h \quad (3.4)$$

(donde en realidad δ es función de h y de x_0 , pero hemos omitido la dependencia de x_0 en la notación para simplificar las fórmulas), vemos que

$$\frac{\delta(h)}{h} = \frac{f(x_0 + h) - f(x_0) - f'(x_0)h}{h} = \frac{f(x_0 + h) - f(x_0)}{h} - f'(x_0)$$

y entonces

$$\lim_{h \rightarrow 0} \frac{\delta(h)}{h} = \lim_{h \rightarrow 0} \frac{f(x_0 + h) - f(x_0)}{h} - f'(x_0) = f'(x_0) - f'(x_0) = 0$$

por la definición de derivada (véase (3.1)). La existencia de ésta, por tanto, nos da, a través de la definición de δ en (3.4), la llamada *fórmula del incremento*

$$f(x_0 + h) - f(x_0) = f'(x_0)h + \delta(h) \quad (3.5)$$

en la cual el incremento $\Delta f(x_0, h)$ de la función aparece descompuesto en suma de un término de primer orden (en h), $f'(x_0)h$, y un resto o *término complementario del incremento*, $\delta(h)$, que es "un infinitésimo de orden superior a h ", terminología un tanto anticuada para expresar que

$$\lim_{h \rightarrow 0} \frac{\delta(h)}{h} = 0 \quad (3.6)$$

(propiedad que también se expresa escribiendo $\delta(h) = o(h)$, que se lee "delta es o pequeña de h cuando $h \rightarrow 0$ ").

Recíprocamente, si se exige que el incremento $\Delta f(x_0, h)$ ha de admitir una descomposición de la forma

$$f(x_0 + h) - f(x_0) = ah + \delta(h) \quad (3.7)$$

con a una cierta constante y $\delta(h)$ verificando (3.6), entonces

$$\frac{f(x_0 + h) - f(x_0)}{h} = a + \frac{\delta(h)}{h} \xrightarrow{h \rightarrow 0} a$$

es decir, f ha de ser derivable en x_0 y $f'(x_0) = a$. Por tanto, las propiedades "f es derivable en x_0 " y " $\Delta f(x_0, h)$ admite una descomposición (3.7) con $\delta(h)$ verificando (3.6)" son equivalentes.

El producto $f'(x_0)h$ es una función de dos variables (x_0 y h) que se llama *diferencial* de f en el punto x_0 correspondiente al incremento h , y se designa $df(x_0; h)$ ó $df(h)$ ó df , según los datos que queramos dar por sobreentendidos:

$$df = df(x_0; h) \stackrel{\text{def}}{=} f'(x_0)h \quad (3.8)$$

donde el distinto papel que juegan las variables x_0 y h justifica que sea usual poner ";" en vez de "y" entre dichas variables. (Obsérvese que, fijado x_0 , la aplicación $h \mapsto df(x_0; h) = f'(x_0)h$ de \mathbb{R} en \mathbb{R} , es lineal).

NOTA. Tradicionalmente, la fórmula del incremento se escribía simplemente $\Delta f = df + \delta$, ó $\Delta y = dy + \delta$, o incluso $\Delta y \simeq dy$, omitiendo ("sobrentendiendo") toda referencia a h (y no digamos a x_0), lo cual ha venido siendo siempre la principal fuente de dificultades en la comprensión de las fórmulas del cálculo diferencial, pues no deja claramente de manifiesto que Δy , dy y δ no son cantidades fijas, sino dependientes de h . En cambio, la "versión completa" (3.5) no es más que una reformulación de la definición de la función δ y no presenta, por tanto, dificultad conceptual ninguna.

La interpretación analítica de (3.5) es la siguiente: la diferencial es la primera aproximación del incremento en el sentido de que el error $\delta(x_0, h) = \delta(h)$ cometido al sustituir $\Delta f(x_0, h)$ por $df(x_0, h)$ tiende a cero más rápidamente que h (o sea, $\delta(h)/h \rightarrow 0$ cuando $h \rightarrow 0$). (En el ejemplo de $f(x) = x^2$, $\Delta f(x_0, h) = 2x_0h + h^2$ y, por tanto, $\delta(h) = h^2$, que evidentemente satisface $\delta(h)/h = h^2/h = h \rightarrow 0$, cuando $h \rightarrow 0$. Análogamente, si $f(x) = x^3$, $\Delta f(x_0, h) = 3x_0^2h + 3x_0h^2 + h^3$, $\delta(h) = \delta(x_0, h) = 3x_0h^2 + h^3$, y $\delta(h)/h = 3x_0h + h^2$. En general, si $f(x)$ es un polinomio, $\delta(h)$ es otro polinomio que empieza por el término de segundo grado, por lo que (3.6) está garantizado: para las funciones polinómicas se puede establecer la fórmula del incremento mediante operaciones algebraicas, más o menos laboriosas pero elementales, sin necesidad de introducir la noción de derivada por vía de un paso al límite; véase la nota 3 de final de capítulo).

La mejor manera de entender el carácter variable de Δy , dy y δ es mediante la fundamental interpretación geométrica de la diferencial, o, si se quiere, de la fórmula del incremento (figura 3.3).

En la figura 3.3, $df = df(x_0, h)$ es el segmento vertical que representa el incremento que, en respuesta al incremento h de la variable independiente, experimentaría la función f si la sustituyéramos por su tangente en el punto x_0 . Por eso se dice que la diferencial no es otra cosa que el incremento medido sobre la tangente, mientras que $\Delta f = \Delta f(x_0, h)$ es el segmento vertical que representa el verdadero incremento de la función, es decir, el medido sobre la curva. La diferencia entre ambos (el "error" $\delta(h)$) puede ser ciertamente grande si h lo es, pero, si nos fijamos en su comportamiento cuando $h \rightarrow 0$, veremos que no sólo tiende a cero con h (cosa que sucedería igualmente si en lugar de la tangente hubiésemos elegido una recta cualquiera que pasase por el punto $(x_0, f(x_0))$ de la curva; hágase la gráfica como ejercicio), sino que, incluso dividido por h , tiende también a 0 cuando $h \rightarrow 0$, siendo esta propiedad, mucho más restrictiva, exclusiva de la tangente —es la propiedad de unicidad de la descomposición (3.5) que acabamos de ver—. Así pues, lo que distingue a la "tangente" (recta que "toca" a la curva) de las "secantes" (rectas que "cortan" a la curva) es que la separación entre curva y recta decrece más rápidamente a cero en la tangente que en cualquiera de las secantes, entendiendo "más rápidamente" en el sentido usual de la comparación por

cociente, como establece con precisión (3.6).

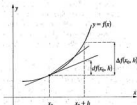


Figura 3.3

En términos de la función $l(x) = y_0 + f'(x_0)(x - x_0)$ definida en (3.2), podemos resumir la discusión geométrica anterior diciendo que l es la única función lineal tal que

$$\frac{f(x) - l(x)}{x - x_0} \rightarrow 0 \quad \text{cuando } x \rightarrow x_0 \quad (3.9)$$

Como vamos a ver, es ésta la vía de generalización más fructífera a las funciones de varias variables; más, incluso, que la propia noción de derivada, por la que, de todos modos, comenzaremos en el siguiente apartado.

3.2 Derivadas parciales

Comenzábamos el apartado anterior recordando que la noción de derivada surge al intentar describir cómo “responde” la variable dependiente y ante variaciones de la variable independiente x en una relación funcional $y = f(x)$. Si ahora tenemos una relación funcional entre varias variables $z = f(x, y)$, que suponemos modela un cierto fenómeno, ¿qué más lógico —en un primer nivel de análisis— que estudiar la variación de z respecto de cada una de las variables x e y , manteniendo fija la otra, como cuando se hace un experimento controlado en el que sólo se altera uno de los factores cada vez? Esta forma de analizar un proceso multivariable tiene mucha tradición en todos los ámbitos de la ciencia, en particular, en el de la ciencia económica, donde se llama *análisis ceteris paribus*, es decir, “siendo iguales todas las demás cosas”. Mediante este artificio nos limitamos al campo de las funciones de una sola variable a través de las *funciones parciales*

$$f(\cdot, y_0) \equiv f^{y_0} : x \mapsto f(x, y_0)$$

$$f(x_0, \cdot) \equiv f^{x_0} : y \mapsto f(x_0, y)$$

que resultan al mantener constante y en el valor y_0 en el primer caso, y x en el valor x_0 en el segundo; podremos entonces utilizar todo el aparato disponible para las funciones de una variable, en particular el cálculo diferencial. Así, partiendo de un punto (x_0, y_0) , los incrementos parciales experimentados por f como respuesta a un incremento de cada una de las variables manteniendo constante la otra son

$$\Delta_x f = f(x_0 + h, y_0) - f(x_0, y_0)$$

$$\Delta_y f = f(x_0, y_0 + k) - f(x_0, y_0)$$

Las correspondientes variaciones relativas o cocientes incrementales parciales son

$$\frac{f(x_0 + h, y_0) - f(x_0, y_0)}{h}$$

$$\frac{f(x_0, y_0 + k) - f(x_0, y_0)}{k}$$

Finalmente, cuando llevamos el análisis al límite, buscando describir la tendencia de la respuesta de z a variaciones de cada variable independiente individual, llegamos al concepto de *derivada parcial*:

Definición 3.1 Sea $f : D \rightarrow \mathbb{R}$ una función definida sobre el subconjunto D de \mathbb{R}^2 , y sea (x_0, y_0) un punto interior de D . Se llama *derivada parcial de f respecto de x en el punto (x_0, y_0)* al límite, si existe,

$$\lim_{h \rightarrow 0} \frac{f(x_0 + h, y_0) - f(x_0, y_0)}{h} \quad (3.10)$$

y se denota por cualquiera de los símbolos siguientes:

$$\frac{\partial f}{\partial x}(x_0, y_0), \left(\frac{\partial f}{\partial x} \right)_{(x_0, y_0)}, f'_x(x_0, y_0), f_x(x_0, y_0), D_x f(x_0, y_0) \text{ ó } D_1 f(x_0, y_0)$$

Análogamente, la *derivada parcial de f respecto de y en el punto (x_0, y_0)* es

$$\lim_{k \rightarrow 0} \frac{f(x_0, y_0 + k) - f(x_0, y_0)}{k} \quad (3.11)$$

y se denota por

$$\frac{\partial f}{\partial y}(x_0, y_0), \left(\frac{\partial f}{\partial y} \right)_{(x_0, y_0)}, f'_y(x_0, y_0), f_y(x_0, y_0), D_y f(x_0, y_0) \text{ ó } D_2 f(x_0, y_0)$$

Si en todo $(x_0, y_0) \in D$ existe $D_x f(x_0, y_0)$, entonces la función de D en \mathbb{R} dada por

$$(x_0, y_0) \mapsto D_x f(x_0, y_0)$$

se llama **función derivada parcial de f respecto de x** , y análogamente se define la **función derivada parcial de f respecto de y** .

Obsérvese que $D_x f(x_0, y_0)$ es la derivada de la función parcial

$$x \mapsto f(x, y_0)$$

evaluada para $x = x_0$. Análogamente, $D_y f(x_0, y_0)$ es la derivada de la función parcial

$$y \mapsto f(x_0, y)$$

particularizada en $y = y_0$. Esto significa que no es necesario introducir un "cálculo especial" para derivadas parciales, sino que la tabla de derivadas tradicional de las funciones de una variable nos es suficiente para calcular derivadas parciales sin dificultad:

EJEMPLO. Sea $f(x, y) = x^2 e^{xy}$. Entonces las funciones derivadas parciales son:

$$\begin{aligned} \frac{\partial f}{\partial x} &= \text{derivada de } x^2 e^{xy} \text{ como función de } x \text{ para } y \text{ fijo} = \\ &= 2x e^{xy} + x^2 (y e^{xy}) = (2x + x^2 y) e^{xy} \\ \frac{\partial f}{\partial y} &= \text{derivada de } x^2 e^{xy} \text{ como función de } y \text{ para } x \text{ fijo} = \\ &= x^2 (x e^{xy}) = x^3 e^{xy} \end{aligned}$$

Las derivadas parciales en el punto $(1, 3)$ serán, pues

$$\begin{aligned} \frac{\partial f}{\partial x}(1, 3) &= [(2x + x^2 y) e^{xy}]_{(x=1, y=3)} = 5e^3 \\ \frac{\partial f}{\partial y}(1, 3) &= [x^3 e^{xy}]_{(x=1, y=3)} = e^3 \end{aligned}$$

A pesar de la ambigüedad que supone (léase la nota 1 de final de capítulo), si representamos por z la variable dependiente, $z = f(x, y)$, también se usan en la práctica las notaciones

$$\frac{\partial z}{\partial x}, \frac{\partial z}{\partial y}, z'_x, z'_y, z_x, z_y$$

para las derivadas parciales.

Interpretación geométrica de las derivadas parciales

La misma propiedad que nos ha permitido incorporar el cálculo explícito de derivadas parciales dentro del marco del cálculo de una variable, es decir, la observación de que las derivadas parciales no son otra cosa que derivadas ordinarias de las funciones de una variable obtenidas fijando todas las variables menos una de ellas, nos va a permitir dar una interpretación geométrica sencilla de aquéllas, apoyándonos en el hecho de que la derivada de una función de una variable en un cierto punto representa la pendiente de la recta tangente en dicho punto a la curva característica de la función.

Fijemos $(x_0, y_0) \in D$. Sabemos que $D_x f(x_0, y_0)$ es la derivada respecto de x de la función $\phi(x) \stackrel{\text{def}}{=} f(x, y_0)$. Observemos ahora que la gráfica de la función ϕ se puede obtener directamente a partir de la superficie $z = f(x, y)$ característica de la función, sin más que particularizar $y = y_0$, lo cual, geoméricamente, corresponde a restringirse al plano vertical $y = y_0$, o sea, a considerar una sección transversal de la superficie. Ésta se llama "curva coordenada en la dirección x que pasa por (x_0, y_0) ", o simplemente, curva x -coordenada, y no es otra cosa que la curva plana obtenida como intersección de la superficie $z = f(x, y)$ con el plano vertical $y = y_0$. Por tanto, la derivada parcial de f respecto de x en (x_0, y_0) es la pendiente de la recta tangente en (x_0, y_0) a la curva coordenada en la dirección x que pasa por dicho punto, recta que llamaremos "tangente (en (x_0, y_0, z_0)) en la dirección x ". Recuérdese la discusión del capítulo 1 para el caso de que $z = f(x, y)$ es un plano, en cuyo caso las curvas coordenadas en la dirección x son simplemente un haz de rectas paralelas. Obsérvese también que, como la curva coordenada está contenida en el plano $y = y_0$, también lo está su tangente en (x_0, y_0, z_0) . Por esta razón el vector tangente en (x_0, y_0, z_0) en la dirección x deberá tener nula su segunda componente, y será, por tanto, $(1, 0, f_x(x_0, y_0))$.

Análogamente, la curva coordenada en la dirección y que pasa por (x_0, y_0) , o curva y -coordenada, sería la curva intersección de la superficie $z = f(x, y)$ con el plano vertical $x = x_0$, y la derivada parcial respecto de y sería la pendiente de la tangente en (x_0, y_0, z_0) a la curva coordenada en la dirección y . El vector tangente a dicha curva es $(0, 1, f_y(x_0, y_0))$.

Estas "curvas coordenadas" definidas sobre la superficie por el procedimiento indicado podrían asimilarse a un sistema de "meridianos y paralelos" sobre la misma, de forma que por cada punto de la superficie pasa exactamente una curva x -coordenada y una curva y -coordenada (fig 3.4).

Crecimiento y decrecimiento parcial

Una de las consecuencias más importantes de la interpretación geométrica de la derivada en funciones de una variable es su aplicabilidad al estudio del creci-

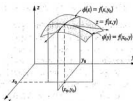


Figura 3.4

miento de una función en un intervalo, a través de su signo. Concretamente, si $f'(x) \geq 0$ para todo x de un intervalo $I = (a, b)$, entonces f es *creciente* en I , y es *estrictamente creciente* si $f'(x) > 0$ en I . Análogamente, si $f'(x) \leq (<) 0$ en I , f es *decreciente* (estrictamente decreciente). Pues bien, es inmediato trasladar esta idea al ámbito de las funciones de varias variables y concluir lo siguiente:

Proposición 3.1 Sea R el rectángulo

$$R = \{(x, y) : a < x < b, c < y < d\} = (a, b) \times (c, d)$$

- Si $\frac{\partial f}{\partial x} \geq 0$ en R , entonces f , como función de x , es *creciente* en (a, b) para todo $y \in (c, d)$. Es decir:

Para todo $y \in (c, d)$ y todo x_1, x_2 con $a < x_1 < x_2 < b$ se tiene:

$$f(x_1, y) \leq f(x_2, y)$$

- Si se tiene la propiedad más fuerte $\frac{\partial f}{\partial x} > 0$ en R , entonces f es *estrictamente creciente* en x , es decir:

Para todo $y \in (c, d)$ y todo x_1, x_2 con $a < x_1 < x_2 < b$ se tiene:

$$f(x_1, y) < f(x_2, y)$$

- Análogamente, si $\frac{\partial f}{\partial x} \leq 0$ (< 0) en R , entonces f es *decreciente* (o *estrictamente decreciente*) como función de x , es decir:

Para todo $y \in (c, d)$ y todo x_1, x_2 con $a < x_1 < x_2 < b$ se tiene:

$$f(x_1, y) \geq f(x_2, y) \text{ (ó } f(x_1, y) > f(x_2, y))$$

Naturalmente, si la derivada parcial con signo constante en R es $\partial f/\partial y$, podemos sacar conclusiones análogas sobre el crecimiento o decrecimiento de f como función de y para cada x constante.

EJEMPLO. Sea la función $f(x, y) = e^{-x^2y}$. Tenemos: $\frac{\partial f}{\partial x} = -2xye^{-x^2y}$,
 $\frac{\partial f}{\partial y} = -x^2e^{-x^2y}$.

- a) Como $\frac{\partial f}{\partial x} < 0$ en el primer cuadrante $\mathbb{R}_{++} = \{x > 0, y > 0\}$ y en el tercer cuadrante $\mathbb{R}_{--} = \{x < 0, y < 0\}$, se tiene:

$$a1) y > 0, 0 < x_1 < x_2 \Rightarrow e^{-x_1^2y} > e^{-x_2^2y}$$

$$a2) y < 0, x_1 < x_2 < 0 \Rightarrow e^{-x_1^2y} > e^{-x_2^2y}$$

- b) Como $\frac{\partial f}{\partial x} > 0$ en el segundo cuadrante $\mathbb{R}_{-+} = \{x < 0, y > 0\}$ y en el cuarto cuadrante $\mathbb{R}_{+-} = \{x > 0, y < 0\}$, se tiene

$$b1) y > 0, x_1 < x_2 < 0 \Rightarrow e^{-x_1^2y} < e^{-x_2^2y}$$

$$b2) y < 0, 0 < x_1 < x_2 \Rightarrow e^{-x_1^2y} < e^{-x_2^2y}$$

- c) Como $\frac{\partial f}{\partial y} \leq 0$ en todo el plano: $x \in \mathbb{R}, y_1 < y_2 \Rightarrow e^{-x^2y_1} \geq e^{-x^2y_2}$

Dejamos como ejercicio al lector perfilar con más detalle esta última desigualdad, teniendo en cuenta que $\partial f/\partial y < 0$ siempre que $x \neq 0$.

También dejamos al lector interpretar geométricamente el contenido de la proposición anterior ($\partial f/\partial x > 0 \Rightarrow$ las pendientes en la dirección del eje Ox son positivas \Rightarrow la superficie $z = f(x, y)$ "crece en la dirección x ", etc.).

Derivadas parciales y continuidad

Una primera diferencia importante con las propiedades de las funciones de una variable es que, a diferencia de éstas, en una función de varias variables, la existencia de derivadas parciales no implica la continuidad. Veámoslo:

EJEMPLO. Consideremos la función

$$f(x, y) = \frac{xy^2}{x^2 + y^4}, \quad (x, y) \neq (0, 0)$$

$$f(0, 0) = 0$$

que estudiamos en el segundo capítulo. A diferencia de los ejemplos anteriores, ahora $f(x, y)$ no viene definida por una única fórmula válida en todos los puntos de su dominio (\mathbb{R}^2 , en este caso). Por ello, para calcular las derivadas parciales en un punto $(x, y) \neq (0, 0)$ podemos aplicar las reglas conocidas de funciones de una variable (pues la función está dada también por $xy^2/(x^2+y^4)$) para puntos suficientemente próximos a él) pero para calcularlas en el origen tenemos que recurrir a la propia definición de derivada parcial, y así:

$$\begin{aligned} D_x f(0, 0) &= \lim_{h \rightarrow 0} \frac{f(0+h, 0) - f(0, 0)}{h} = \lim_{h \rightarrow 0} \frac{\frac{h \cdot 0}{h^2+0} - 0}{h} = \lim_{h \rightarrow 0} \frac{0-0}{h} = 0 \\ D_y f(0, 0) &= \lim_{k \rightarrow 0} \frac{f(0, 0+k) - f(0, 0)}{k} = \lim_{k \rightarrow 0} \frac{\frac{0 \cdot k^2}{0+k^4} - 0}{k} = \lim_{k \rightarrow 0} \frac{0-0}{k} = 0 \end{aligned}$$

Existen, por tanto, las dos derivadas parciales en el origen; sin embargo, f no es continua en $(0, 0)$, como se vio entonces, a pesar de que existen los límites de f en $(0, 0)$ según rectas y tienen todos el mismo valor $f(0, 0) = 0$.

Se rompe así la escala de regularidad: continuidad \rightarrow derivabilidad que se tenía para las funciones de una variable. Por supuesto, la existencia de derivadas parciales de $f(x, y)$ implica la continuidad de las correspondientes funciones parciales f^y, f^x (o sea, la continuidad de f en (x_0, y_0) según las rectas $y = y_0, x = x_0$) pero la continuidad de $f(x, y)$ como función de dos variables es una propiedad más fuerte que eso, como sabemos. Se ve así la necesidad de introducir una noción más restrictiva que implique la continuidad. Ésta será la *diferenciabilidad*, que trataremos en detalle en el apartado que sigue y que, como apuntábamos al final del apartado anterior, nos va a permitir una descripción del efecto *combinado* que produce en f una *variación simultánea* de todas las variables de las que depende.

Funciones de n variables

La extensión de la definición 1 a funciones de n variables es inmediata (recordemos que denotamos e_i o \tilde{e}_i a los vectores de la base canónica de \mathbb{R}^n y simplemente por 0 al vector $(0, \dots, 0)$ de \mathbb{R}^n):

Definición 3.2 Sea $f: D \rightarrow \mathbb{R}$, donde $D \subset \mathbb{R}^n$, y sea $x_0 = (x_1^0, x_2^0, \dots, x_n^0)$ un punto interior de D . La derivada parcial de f respecto de x_i en x_0 es el límite, si existe

$$\lim_{h \rightarrow 0} \frac{f(x_1^0, \dots, x_{i-1}^0, x_i^0 + h, x_{i+1}^0, \dots, x_n^0) - f(x_1^0, \dots, x_{i-1}^0, x_i^0, x_{i+1}^0, \dots, x_n^0)}{h} \quad (3.12)$$

es decir,

$$\lim_{h \rightarrow 0} \frac{f(x_0 + he_i) - f(x_0)}{h}$$

y se denota por cualquiera de los símbolos

$$D_{x_i}f(x_0), \partial_{x_i}f(x_0), D_i f(x_0), \partial_i f(x_0), \frac{\partial f}{\partial x_i}(x_0), \left(\frac{\partial f}{\partial x_i}\right)_{x_0}, f_{x_i}(x_0), f'_{x_i}(x_0)$$

Si en todo punto $x_0 \in D$ existe $D_{x_i}f(x_0)$, entonces la función de D en \mathbb{R} dada por

$$x_0 \mapsto D_{x_i}f(x_0)$$

se llama **función derivada parcial de f respecto de x_i** .

EJEMPLO. Sea $f(x, y, z) = e^{xyz} \cos \frac{yz}{x}$. Se tiene:

$$\frac{\partial f}{\partial x} = yze^{xyz} \cos \frac{yz}{x} + e^{xyz} \left(\operatorname{sen} \frac{yz}{x} \right) \frac{yz}{x^2}$$

$$\frac{\partial f}{\partial y} = xze^{xyz} \cos \frac{yz}{x} - e^{xyz} \left(\operatorname{sen} \frac{yz}{x} \right) \frac{z}{x}$$

$$\frac{\partial f}{\partial z} = xye^{xyz} \cos \frac{yz}{x} - e^{xyz} \left(\operatorname{sen} \frac{yz}{x} \right) \frac{y}{x}$$

3.3 Diferenciabilidad y aproximación lineal

El análisis parcial o *ceteris paribus* de una cierta relación funcional multivariable que nos ha llevado al concepto de derivada parcial tiene, sin duda, un valor importante, pero, como su propio nombre indica, no parece claro a primera vista cómo podría proporcionarnos información sobre el efecto combinado que produciría una *variación simultánea* de todas las variables. Veamos bajo qué condiciones ello puede ser posible. Recordemos del primer capítulo que si en una *función lineal*

$$f(x, y) = ax + by + c$$

x se incrementa en h unidades e y en k unidades, el incremento global de f es

$$\Delta f = f(x_0 + h, y_0 + k) - f(x_0, y_0) = ah + bk$$

que no es otra cosa que la suma de los incrementos individuales (parciales) generados por las variaciones de x para y constante (es decir, cuando $k = 0$) y la de y para x constante ($h = 0$). En otras palabras, podemos *descomponer* el incremento total en sus incrementos parciales y considerar el proceso global como una *superposición* de acciones: primero incrementamos x_0 sin variar y hasta llegar a $x_0 + h$ y luego, sin modificar este valor, incrementamos y_0 hasta alcanzar el valor $y_0 + k$. Así, el proceso incremental total no es más que la suma de sus partes.

Para funciones arbitrarias, la situación ya no es la misma. Consideremos la extensión natural del ejemplo del cuadrado que veíamos en la introducción, llamando ahora $f(x, y) = xy$ al área de un rectángulo de lados x e y , considerados como dos variables independientes. Demos incrementos h y k a x_0 e y_0 , respectivamente, de forma que el área del rectángulo de lados $x_0 + h$ e $y_0 + k$ será (figura 3.5):

$$f(x_0 + h, y_0 + k) = (x_0 + h)(y_0 + k) = x_0 y_0 + y_0 h + x_0 k + hk$$

El incremento de área asociado a los incrementos h y k de los lados es

$$\begin{aligned}\Delta f &= \Delta f(x_0, y_0; h, k) \stackrel{\text{def}}{=} f(x_0 + h, y_0 + k) - f(x_0, y_0) = \\ &= (x_0 + h)(y_0 + k) - x_0 y_0 = y_0 h + x_0 k + hk\end{aligned}$$

descomposición cuya interpretación geométrica es evidente a la partir de la figura 3.5 (compárese con la figura 3.2).

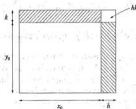


Figura 3.5

Por otra parte, los incrementos parciales, obtenidos fijando uno de los lados, son

$$\begin{aligned}\Delta_x f &= f(x_0 + h, y) - f(x_0, y) = y_0 h, \\ \Delta_y f &= f(x_0, y_0 + k) - f(x_0, y_0) = x_0 k\end{aligned}$$

con lo que

$$\Delta f = \Delta_x f + \Delta_y f + hk \neq \Delta_x f + \Delta_y f$$

es decir, ahora el proceso incremental global no se puede obtener como superposición de los procesos incrementales parciales. Pero, como en el caso de las funciones de una variable, cabe esperar que fórmulas y procesos válidos en el ámbito de la linealidad que dejan de serlo en cuanto la abandonamos, puedan servir de *aproximación* al proceso no lineal (en este caso, que lo que no es exacto a nivel de incremento pueda serlo a nivel de la diferencial), a condición de que aclaremos perfectamente qué entendemos por aproximación. Desde el punto de vista práctico, la clave está, como resaltábamos en la introducción, en la metodología general de descomponer los incrementos en *órdenes de magnitud* claramente diferenciados. En el ejemplo del área del rectángulo $f(x, y) = xy$, y partiendo de unos valores iniciales $x = x_0, y = y_0$, analicemos los tamaños relativos de los elementos de la fórmula de Δf : dado que h y k son incrementos independientes, no podemos singularizar *a priori* ninguno de los términos $y_0 h$ ó $x_0 k$, que representan las áreas de los rectángulos laterales de la figura 3.5, o sea, hay que considerarlos conjuntamente; sin embargo, si tanto h como k son pequeños, de lo que no hay duda es de que el producto hk es mucho más pequeño que los otros dos términos, lo cual se corresponde con el menor tamaño relativo del rectángulo del ángulo superior derecho en comparación con los rectángulos laterales antes mencionados. Dicho de otra forma, la *descomposición en escalas* de magnitud del incremento Δf en términos de los incrementos h y k debe ser la siguiente:

$$\Delta f = \underbrace{y_0 h + x_0 k}_{\text{primer orden}} + \underbrace{hk}_{\text{segundo orden}}$$

Si se tratase, por ejemplo, de la función $f(x, y) = x^2 y + y^3$, se obtendría

$$\begin{aligned} \Delta f &= \{(x_0 + h)^2 (y_0 + k) + (y_0 + k)^3\} - \{x_0^2 y_0 + y_0^3\} = \\ &= \underbrace{\{2x_0 y_0 h + (x_0^2 + 3y_0^2)k\}}_{\text{primer orden}} + \underbrace{\{y_0 h^2 + 3y_0 k^2 + 2x_0 h k\}}_{\text{segundo orden}} + \underbrace{\{h^2 k + k^3\}}_{\text{tercer orden}} \end{aligned}$$

y análogamente para cualquier polinomio en x, y . En todos estos casos encontramos una expresión (o *fórmula del incremento*) del tipo

$$\Delta f = f(x_0 + h, y_0 + k) - f(x_0, y_0) = \underbrace{\{ah + bk\}}_{\text{primer orden}} + \underbrace{g(h, k)}_{\text{orden superior}} \quad (3.13)$$

con unos ciertos coeficientes a, b , en la que Δf aparece como suma de un término de primer orden (es decir, *lineal*) en (h, k) más un *resto* o *término*

complementario $g(h, k)$ cuya mayor "pequeñez" para h y k pequeños queda exactamente reflejada en el hecho de que es un infinitésimo de orden superior a la longitud del vector de incrementos (h, k) , es decir, que verifica

$$\frac{g(h, k)}{|(h, k)|} \rightarrow 0 \text{ cuando } |(h, k)| = \sqrt{h^2 + k^2} \rightarrow 0 \quad (3.14)$$

lo que se comprueba fácilmente en estos ejemplos. También se comprueba inmediatamente para ellos que los coeficientes a y b son precisamente $f'_x(x_0, y_0)$ y $f'_y(x_0, y_0)$, con lo que (3.13) tiene la forma

$$\Delta f = \frac{\partial f}{\partial x}(x_0, y_0)h + \frac{\partial f}{\partial y}(x_0, y_0)k + g(h, k) \quad (3.15)$$

con $g(h, k)$ verificando (3.14), fórmula de descomposición del incremento homóloga a la fórmula (3.5) obtenida para el área del cuadrado y para cualquier otra función de una variable derivable (utilizando ahora, como es lógico, la longitud $\sqrt{h^2 + k^2}$ del vector plano de incrementos en lugar de la longitud $|h|$ sobre la recta para medir el grado de aproximación de la *parte lineal* respecto del incremento total). ¿Se tendrá también la *fórmula del incremento* (3.15) para una función arbitraria $f(x, y)$ que posea derivadas parciales en el punto (x_0, y_0) ? En otras palabras: contando con que los incrementos parciales, relativos a funciones de una variable, admiten la fórmula de aproximación

$$\Delta_x f = f(x_0 + h, y_0) - f(x_0, y_0) = \frac{\partial f}{\partial x}(x_0, y_0)h + \delta_1(h)$$

$$\Delta_y f = f(x_0, y_0 + k) - f(x_0, y_0) = \frac{\partial f}{\partial y}(x_0, y_0)k + \delta_2(k)$$

donde δ_1 y δ_2 son "infinitésimos de orden superior" a h y k , respectivamente ¿podría ser que en la fórmula

$$\begin{aligned} \Delta f &= f(x_0 + h, y_0 + k) - f(x_0, y_0) = \Delta_x f + \Delta_y f + \delta_3(h, k) = \\ &= \frac{\partial f}{\partial x}(x_0, y_0)h + \frac{\partial f}{\partial y}(x_0, y_0)k + g(h, k) \end{aligned}$$

donde $g = \delta_1 + \delta_2 + \delta_3$, ocurriese que $g(h, k)$ fuese, a su vez, un infinitésimo de orden superior al incremento global (h, k) , entendiendo este concepto en el sentido de (3.14)? Pues bien, la respuesta es no: a diferencia de las funciones de una variable, la mera existencia de derivadas parciales de una función de más de una variable no garantiza la validez de la *fórmula del incremento* (3.15). Veamos un contraejemplo debido a THOMAE. Sea

$$f(x, y) = \begin{cases} \frac{xy}{\sqrt{x^2 + y^2}} & \text{si } (x, y) \neq (0, 0) \\ 0 & \text{si } (x, y) = (0, 0) \end{cases}$$

Se tiene $D_x f(0,0) = D_y f(0,0) = 0$, mientras que

$$\begin{aligned}\frac{g(h,k)}{\sqrt{h^2+k^2}} &= \frac{1}{\sqrt{h^2+k^2}} \left[f(0+h,0+k) - f(0,0) - \frac{\partial f}{\partial x}(0,0)h - \frac{\partial f}{\partial y}(0,0)k \right] \\ &= \frac{1}{\sqrt{h^2+k^2}} \frac{hk}{\sqrt{h^2+k^2}} = \frac{hk}{h^2+k^2}\end{aligned}$$

expresión que carece de límite cuando $(h,k) \rightarrow (0,0)$, como se comprueba fácilmente viendo que los límites según las rectas $k = mh$ dependen de m . Por otra parte, la función f sí es continua en $(0,0)$ (demuéstrese).

Pero, como quedaba apuntado al final de la Introducción, la generalización realmente fructífera a funciones de varias variables es la del hecho de que el concepto de derivada permite, en las funciones de una variable, una aproximación local (en el entorno de un punto x_0), "salvo infinitésimos de orden superior", de una función cualquiera no lineal por medio de una función lineal, es decir, lo que interesa es que se mantenga la validez de la fórmula del incremento. Como esto no está garantizado para funciones derivables (que poseen derivadas parciales) se hace necesario, para progresar en el cálculo diferencial de funciones de varias variables, introducir un concepto más restringido, el de *función diferenciable*, que es, simplemente, aquella para la que es válida la fórmula del incremento:

Definición 3.3 Sea f una función definida en un conjunto $D \subset \mathbb{R}^2$, y sea (x_0, y_0) un punto interior de D . Se dice que f es **diferenciable** en (x_0, y_0) si existen $f'_x(x_0, y_0)$, $f'_y(x_0, y_0)$ y, para todo h, k suficientemente pequeños, se tiene:

$$f(x_0 + h, y_0 + k) - f(x_0, y_0) = f'_x(x_0, y_0)h + f'_y(x_0, y_0)k + g(h, k) \quad (3.16)$$

donde

$$g(h, k) \stackrel{\text{def}}{=} f(x_0 + h, y_0 + k) - f(x_0, y_0) - \{f'_x(x_0, y_0)h + f'_y(x_0, y_0)k\}$$

llamada *resto, término de error o término complementario del incremento*, es una expresión que satisface

$$\lim_{(h,k) \rightarrow (0,0)} \frac{g(h,k)}{\sqrt{h^2+k^2}} = 0 \quad (3.17)$$

La cantidad

$$df(x_0, y_0; h, k) \stackrel{\text{def}}{=} \frac{\partial f}{\partial x}(x_0, y_0)h + \frac{\partial f}{\partial y}(x_0, y_0)k \quad (3.18)$$

considerada como *expresión lineal* en (h, k) , se llama *diferencial de f en (x_0, y_0)* . Se dice que f es *diferenciable en un conjunto abierto D* si lo es en cada punto de D .

La cuestión de la diferenciabilidad es delicada en términos teóricos, aunque, en la práctica, podremos garantizar que la mayoría de las funciones de uso habitual tienen, como $f(x, y) = xy$ y $f(x, y) = x^2y + y^3$, esta propiedad. En efecto, se tiene el importante resultado siguiente:

Teorema 3.1 *Si existen las derivadas parciales de f en todo un entorno U de (x_0, y_0) y son funciones continuas en (x_0, y_0) , entonces f es diferenciable en (x_0, y_0) .*

La demostración de este teorema es de cierta dificultad y la posponemos al final de este apartado. La condición del enunciado *no es necesaria* para asegurar la diferenciabilidad; es decir, f puede ser diferenciable en (x_0, y_0) , tener derivadas definidas en todo un entorno U de (x_0, y_0) , y ser, incluso ambas, discontinuas en (x_0, y_0) : la función de una variable

$$\varphi(x) = \begin{cases} x^2 \operatorname{sen} \frac{1}{x} & \text{si } x \neq 0 \\ 0 & \text{si } x = 0 \end{cases}$$

es un ejemplo bien conocido de función derivable en todo $x \in \mathbb{R}$ —y, por tanto, diferenciable, en el sentido de que, al tratarse de una función de una variable, es válida la fórmula del incremento (3.5)—, cuya derivada $\varphi'(x)$ es una función discontinua en $x_0 = 0$ (compruébese como ejercicio). De ello se deduce inmediatamente que la función de dos variables

$$f(x, y) = \varphi(x) + \varphi(y)$$

es diferenciable en el origen $(0, 0)$ y tiene derivadas parciales en todo punto $(x, y) \in \mathbb{R}^2$ que no son continuas en $(0, 0)$ (complétense los detalles como ejercicio).

A pesar de no ser en rigor necesaria, la condición dada por el teorema es “la” condición de diferenciabilidad a todos los efectos prácticos, es decir, en lo que respecta a las funciones que normalmente aparecen en las aplicaciones. Esto justifica la siguiente

Definición 3.4 *Una función $f : D$ (abierto) $\rightarrow \mathbb{R}$ es diferenciable con continuidad o continuamente diferenciable en D si existen las derivadas parciales $\partial f / \partial x$, $\partial f / \partial y$ y son funciones continuas en todo D . Se denota por $C^1(D)$ al conjunto de todas las funciones continuamente diferenciables sobre D , que, por esta razón, se llaman también **funciones de clase uno** en D .*

NOTA. Si se verifica esa condición, f es diferenciable en todos los puntos de D y se tiene que

$$df(x, y; h, k) = \frac{\partial f}{\partial x}(x, y)h + \frac{\partial f}{\partial y}(x, y)k$$

es una función continua en las cuatro variables (x, y, h, k) . La posibilidad de que sea, además, diferenciable en (x, y) (en (h, k) lo es, obviamente) depende de que las derivadas parciales $\partial f/\partial x$, $\partial f/\partial y$ sean, a su vez, diferenciables, lo cual nos lleva al tema de las derivadas sucesivas de una función, que estudiaremos en el capítulo 5.

Unicidad de la diferencial

Toda aplicación lineal de \mathbb{R}^2 en \mathbb{R} es de la forma

$$(h, k) \longrightarrow ah + bk$$

para unos ciertos números reales a y b . Pues bien, como para las funciones de una variable, la diferencial

$$\frac{\partial f}{\partial x}(x_0, y_0)h + \frac{\partial f}{\partial y}(x_0, y_0)k$$

es la única expresión lineal en (h, k) que aproxima al incremento $f(x_0 + h, y_0 + k) - f(x_0, y_0)$ salvo "infinitésimos de orden superior", es decir, salvo términos de error $g(h, k)$ que satisfacen (3.17). En efecto, supongamos que existen a y b reales tales que

$$f(x_0 + h, y_0 + k) - f(x_0, y_0) = ah + bk + g(h, k) \quad (3.19)$$

con $g(h, k)$ satisfaciendo (3.17). Se trata de ver que, entonces, necesariamente se tiene $a = \frac{\partial f}{\partial x}(x_0, y_0)$, $b = \frac{\partial f}{\partial y}(x_0, y_0)$: si particularizamos (3.19) para el caso $k = 0$, obtenemos:

$$f(x_0 + h, y_0) - f(x_0, y_0) = ah + g(h, 0)$$

de donde

$$\frac{f(x_0 + h, y_0) - f(x_0, y_0)}{h} = a + \frac{g(h, 0)}{h} \quad (3.20)$$

Vamos a probar que $g(h, 0)/h$ tiene límite 0 cuando $h \rightarrow 0$. Sea $\varepsilon > 0$. Por la hipótesis (3.17), existe $\delta > 0$ tal que

$$0 < \sqrt{h^2 + k^2} < \delta \Rightarrow \left| \frac{g(h, k)}{\sqrt{h^2 + k^2}} \right| < \varepsilon$$

En particular, si $k = 0$ y $0 < |h| < \delta$, entonces $|(h, 0)| = \sqrt{h^2 + 0} < \delta$ y tenemos

$$\left| \frac{g(h, 0)}{h} \right| = \frac{|g(h, 0)|}{|h|} = \left| \frac{g(h, 0)}{|(h, 0)|} \right| < \varepsilon$$

Por tanto, dado $\varepsilon > 0$, el mismo $\delta = \delta(\varepsilon)$ cuya existencia impone la definición de límite para la función $g(h, k)$ de dos variables, nos sirve como valor que garantiza que, si $0 < |h| < \delta$, se tiene

$$\left| \frac{g(h, 0)}{h} \right| < \varepsilon$$

Por definición de límite para funciones de una variable, esto significa que

$$\lim_{h \rightarrow 0} \frac{g(h, 0)}{h} = 0$$

como queríamos demostrar.

Podemos ahora tomar el límite cuando $h \rightarrow 0$ en ambos miembros de (3.20), obteniendo

$$a = \lim_{h \rightarrow 0} \frac{f(x_0 + h, y_0) - f(x_0, y_0)}{h}$$

que no es otra cosa que la derivada parcial de f respecto de x en el punto (x_0, y_0) . Análogamente se demuestra que b es la derivada parcial de f respecto de y en dicho punto.

De acuerdo con esto, podríamos haber adoptado como definición de función diferenciable un enunciado en principio más general que el de la definición 3.3:

Definición 3.5 Sea f una función definida en un conjunto $D \subset \mathbb{R}^2$, y sea (x_0, y_0) un punto interior de D . Se dice que f es **diferenciable** en (x_0, y_0) si existen a, b reales tales que, para todo h, k suficientemente pequeños, se tiene:

$$f(x_0 + h, y_0 + k) - f(x_0, y_0) = ah + bk + g(h, k)$$

donde $g(h, k)$, llamada **resto**, **término de error** o **término complementario del incremento**, es una expresión que satisface

$$\lim_{(h,k) \rightarrow (0,0)} \frac{g(h, k)}{\sqrt{h^2 + k^2}} = 0$$

La cantidad

$$df(x_0, y_0; h, k) \stackrel{\text{def}}{=} ah + bk$$

considerada como **expresión lineal** en (h, k) , se llama **diferencial de f en (x_0, y_0)** , y los números a y b se llaman **coeficientes diferenciales**.

Se dice que f es **diferenciable en un conjunto abierto D** si lo es en cada punto de D .

Pero el argumento anterior implica que, de hecho, si f es diferenciable en (x_0, y_0) según esta definición aparentemente más general, entonces, automáticamente, tiene derivadas parciales en (x_0, y_0) y éstas coinciden con los coeficientes diferenciales a y b , de modo que la diferencial de f en (x_0, y_0) se expresa de la forma (3.18).

NOTAS. 1. La condición (3.17) suele expresarse también omitiendo la referencia directa al resto, simplemente escribiendo

$$\lim_{(h,k) \rightarrow (0,0)} \frac{f(x_0 + h, y_0 + k) - f(x_0, y_0) - \{f'_x(x_0, y_0)h + f'_y(x_0, y_0)k\}}{\sqrt{h^2 + k^2}} = 0$$

o también

$$\lim_{(x,y) \rightarrow (x_0,y_0)} \frac{f(x,y) - f(x_0,y_0) - \{f'_x(x_0,y_0)(x-x_0) + f'_y(x_0,y_0)(y-y_0)\}}{\sqrt{(x-x_0)^2 + (y-y_0)^2}} = 0$$

Poniendo

$$l(x,y) = f(x_0, y_0) + f'_x(x_0, y_0)(x-x_0) + f'_y(x_0, y_0)(y-y_0)$$

vemos que la propiedad de ser f diferenciable en (x_0, y_0) significa que l es la única función lineal tal que

$$\lim_{(x,y) \rightarrow (x_0,y_0)} \frac{f(x,y) - l(x,y)}{\sqrt{(x-x_0)^2 + (y-y_0)^2}} = 0$$

o sea, la única función lineal que aproxima a f cerca de (x_0, y_0) salvo infinitésimos de orden superior al módulo $\sqrt{(x-x_0)^2 + (y-y_0)^2}$ del vector incremento de las variables independientes.

2. Como la función $g(h,k) \stackrel{\text{def}}{=} f(x_0 + h, y_0 + k) - \{f'_x(x_0, y_0)h + f'_y(x_0, y_0)k\}$ de (3.16) está definida para todo (h,k) tal que el punto $(x_0 + h, y_0 + k)$ pertenezca a D , y (x_0, y_0) es un punto interior de D , o sea, existe $\delta > 0$ tal que $B_\delta((x_0, y_0))$ está contenido en D , vemos que la mencionada condición se verifica para todo (h,k) cuya norma $\sqrt{h^2 + k^2}$ sea menor que δ . Por tanto, tiene sentido plantearse la existencia de límite de la función "error relativo" $g(h,k)/\sqrt{h^2 + k^2}$ cuando (h,k) tiende a $(0,0)$.

3. Si (x_0, y_0) no fuese punto interior, sino punto frontera de D , cabría aún considerar dicho límite según la definición dada en el capítulo 2. Esta idea se correspondería con el uso de la noción de derivada lateral para definir la derivada de una función de una variable en los extremos de un intervalo cerrado; pero mientras que en el caso unidimensional esto no encierra ninguna dificultad, para funciones de varias variables esta extensión tropieza con cuestiones mucho más delicadas, razón por la cual no la consideraremos aquí, y nos limitaremos a considerar la noción de diferenciability exclusivamente para puntos interiores. Al decir que f es diferenciable en un conjunto $A \subset \mathbb{R}^2$ se entiende que lo es en un conjunto abierto D que contiene a A .

4. Recordemos que la propiedad (3.17) se expresa escribiendo $g(h,k) = o(|(h,k)|)$ cuando $|(h,k)| \rightarrow 0$ —que se lee "g es o pequeña de $|(h,k)|$ cuando $|(h,k)| \rightarrow 0$ — y también diciendo que " $g(h,k)$ es un infinitésimo de orden superior a la norma $|(h,k)|$ del vector de incrementos".

5. Explicitando la definición de límite, la propiedad (3.17) se puede enunciar del modo equivalente: dado $\varepsilon > 0$, existe $\delta > 0$ tal que

$$|g(h, k)| \leq \varepsilon \sqrt{h^2 + k^2} \text{ para todo } (h, k) \text{ que verifica } \sqrt{h^2 + k^2} \leq \delta. \quad (3.21)$$

(Obsérvese que 3.21 se verifica obviamente para $(h, k) = (0, 0)$ pues $g(0, 0) = 0$ por (3.16)).

6. El punto de vista de la definición 3.5 es interesante para el concepto de diferenciabilidad de funciones vectoriales y en otros contextos todavía más generales.

7. Volviendo al comentario inicial de este apartado, y para que quede perfectamente clara la importancia de la diferenciabilidad, supongamos que lo único que está a nuestro alcance en el estudio de una cierta relación funcional multivariable es su análisis parcial o *ceteris paribus*, y, por tanto, conocemos sus derivadas parciales, pero nada más. Como decíamos, éstas tienen, sin duda, su importancia, pero, por sí solas, no pueden proporcionarnos información sobre el efecto conjunto que produce una variación simultánea de todas las variables, pues, como hemos visto, el incremento total no es la suma de los incrementos parciales. Sin embargo, a la luz de la fórmula del incremento (3.16), podemos afirmar que, en el primer nivel de aproximación, el análisis *ceteris paribus* es útil siempre que la función sea diferenciable. Esto nos indica que, si ya en el cálculo de funciones de una variable la correspondiente fórmula del incremento (3.5) es importante, la expresión multivariable análoga $\Delta f = f'_x(x_0, y_0)h + f'_y(x_0, y_0)k + \dots$ tiene aún más trascendencia, pues no sólo supone una simplificación importante en los cálculos incrementales, sino que justifica el análisis individualizado variable por variable. Esta observación decisiva es la clave de la aplicabilidad de los métodos diferenciales en disciplinas que, como la Economía, manejan sistemáticamente magnitudes dependientes de varias variables y basan gran parte de sus hipótesis y desarrollo en el análisis *ceteris paribus*.

8. Compruébese como ejercicio que si f, g son funciones diferenciables en x_0 , entonces las funciones $f + g$, αf ($\alpha \in \mathbb{R}$) y fg son también diferenciables en x_0 .

Expresión equivalente del resto

Algunos autores prefieren descomponer el resto que aparece en la definición de diferencial de la forma equivalente siguiente:

$$g(h, k) = h\delta_1(h, k) + k\delta_2(h, k) \quad (3.22)$$

donde $\lim_{(h,k) \rightarrow (0,0)} \delta_1(h, k) = \lim_{(h,k) \rightarrow (0,0)} \delta_2(h, k) = 0$. Es fácil ver que si $g(h, k)$ se puede expresar así, forzosamente satisface

$$\lim_{(h,k) \rightarrow (0,0)} \frac{g(h, k)}{\sqrt{h^2 + k^2}} = 0$$

Recíprocamente, si $g(h, k)$ satisface esta última propiedad basta descomponer:

$$\begin{aligned} g(h, k) &= \frac{h^2 + k^2}{\sqrt{h^2 + k^2} \sqrt{h^2 + k^2}} g(h, k) = \\ &= h \left(\frac{h}{\sqrt{h^2 + k^2}} \frac{g(h, k)}{\sqrt{h^2 + k^2}} \right) + k \left(\frac{k}{\sqrt{h^2 + k^2}} \frac{g(h, k)}{\sqrt{h^2 + k^2}} \right) \stackrel{\text{def}}{=} \\ &= h\delta_1(h, k) + k\delta_2(h, k) \end{aligned}$$

y observar que $h/\sqrt{h^2 + k^2}$ y $k/\sqrt{h^2 + k^2}$ están acotados (entre -1 y 1) mientras que $g(h, k)/\sqrt{h^2 + k^2}$ tiende a 0 por hipótesis. Como veremos, esta separación del resto en sumandos que contienen a los incrementos como factores aligera bastante algunas demostraciones, pero en otras es más conveniente tratar al resto como un sólo bloque.

Con esta expresión para el resto, se puede reformular la noción de diferenciabilidad así:

Definición 3.6 f es diferenciable en el punto (x_0, y_0) si el incremento $f(x_0 + h, y_0 + k) - f(x_0, y_0)$ se puede expresar en la forma

$$\begin{aligned} f(x_0 + h, y_0 + k) - f(x_0, y_0) &= \frac{\partial f}{\partial x}(x_0, y_0)h + \frac{\partial f}{\partial y}(x_0, y_0)k + \\ &\quad + h\delta_1(h, k) + k\delta_2(h, k) \end{aligned} \quad (3.23)$$

donde $\delta_1(h, k)$, $\delta_2(h, k)$ son funciones tales que

$$\lim_{(h,k) \rightarrow (0,0)} \delta_1(h, k) = \lim_{(h,k) \rightarrow (0,0)} \delta_2(h, k) = 0 \quad (3.24)$$

Volveremos a encontrar una multiplicidad parecida —o aún mayor— de expresiones para el “término complementario” en el apartado dedicado a la fórmula de Taylor.

Diferenciabilidad y continuidad

Una primera consecuencia inmediata de la propia definición de diferenciabilidad es la *continuidad* de la función, con lo cual se arregla la “anomalía” acerca de las derivadas parciales que vimos en el apartado anterior.

Proposición 3.2 *Toda función diferenciable en un punto es continua en él.*

Demostración: Utilizando (3.23) se tiene, por las propiedades (3.24) de las funciones $\delta_1(h, k)$, $\delta_2(h, k)$,

$$\begin{aligned} & \lim_{(h,k) \rightarrow (0,0)} \{f(x_0 + h, y_0 + k) - f(x_0, y_0)\} = \\ &= \frac{\partial f}{\partial x}(x_0, y_0) \cdot \lim_{(h,k) \rightarrow (0,0)} h + \frac{\partial f}{\partial y}(x_0, y_0) \cdot \lim_{(h,k) \rightarrow (0,0)} k + \\ &+ \lim_{(h,k) \rightarrow (0,0)} h \cdot \lim_{(h,k) \rightarrow (0,0)} \delta_1(h, k) + \lim_{(h,k) \rightarrow (0,0)} k \cdot \lim_{(h,k) \rightarrow (0,0)} \delta_2(h, k) = 0 \end{aligned}$$

o sea

$$\lim_{(x,y) \rightarrow (x_0, y_0)} \{f(x, y) - f(x_0, y_0)\} = 0$$

es decir

$$\lim_{(x,y) \rightarrow (x_0, y_0)} f(x, y) = f(x_0, y_0)$$

que es la propia definición de continuidad en (x_0, y_0) .

NOTA. Como ya sucedía en las funciones de una variable, el recíproco no es cierto, es decir, la continuidad no implica la diferenciabilidad, como muestra el contraejemplo de THOMAE (introdutor del concepto de diferenciabilidad) que vimos anteriormente.

Interpretación geométrica de la diferencial: el plano tangente

Como hemos visto, la fórmula que define la diferencial escrita en la forma

$$\begin{aligned} f(x, y) &= f(x_0, y_0) + f'_x(x_0, y_0)(x - x_0) + f'_y(x_0, y_0)(y - y_0) + \\ &+ g(x - x_0, y - y_0) \end{aligned} \quad (3.25)$$

puede interpretarse de la siguiente forma: Sean $z_0 = f(x_0, y_0)$ y

$$l(x, y) \stackrel{\text{def}}{=} z_0 + f'_x(x_0, y_0)(x - x_0) + f'_y(x_0, y_0)(y - y_0)$$

Entonces $l(x, y)$ es la única expresión lineal en x, y que aproxima a $f(x, y)$ salvo infinitésimos de orden superior al módulo $\sqrt{(x - x_0)^2 + (y - y_0)^2}$ del vector incremento de las variables independientes. Como sabemos del primer capítulo, $Z = l(x, y)$ es la ecuación del plano que pasa por (x_0, y_0, z_0) cuyas pendientes en las direcciones x e y son $f'_x(x_0, y_0)$ y $f'_y(x_0, y_0)$, respectivamente, o, en otros términos, el plano que pasa por (x_0, y_0, z_0) y tiene por vectores directores

$$(1, 0, f'_x(x_0, y_0)) \quad \text{y} \quad (0, 1, f'_y(x_0, y_0)) \quad (3.26)$$

Contiene, por tanto, a las que hemos denominado, al dar en el apartado 3.2 la interpretación geométrica de las derivadas parciales, *rectas tangentes* en (x_0, y_0, z_0) a la superficie $z = f(x, y)$ en las direcciones x e y , y, por ello, parece lógico llamarlo **plano tangente** en (x_0, y_0, z_0) a la superficie $z = f(x, y)$ (nombre que se verá "reforzado" cuando veamos en el capítulo siguiente que, de hecho, contiene a la recta tangente en (x_0, y_0, z_0) en cualquier dirección).

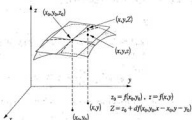


Figura 3.6

Podemos entonces interpretar (3.25) de la siguiente forma: *La diferencial*

$$df(x_0, y_0; x - x_0, y - y_0) = f'_x(x_0, y_0)(x - x_0) + f'_y(x_0, y_0)(y - y_0) \quad (3.27)$$

representa el incremento que experimentarí la función al pasar de (x_0, y_0) a (x, y) , si ésta se sustituyese por su aproximación lineal en (x_0, y_0) . En otras palabras, es el incremento medido sobre el plano tangente, a diferencia del incremento real $f(x, y) - f(x_0, y_0)$, que se mide sobre la superficie $z = f(x, y)$. (Figura 3.6.)

Explicitando, se tiene que la ecuación del plano tangente en (x_0, y_0, z_0) es

$$Z - z_0 = \frac{\partial f}{\partial x}(x_0, y_0)(x - x_0) + \frac{\partial f}{\partial y}(x_0, y_0)(y - y_0) \quad (3.28)$$

Tenemos así una analogía completa con el caso de las funciones de una variable y su interpretación de la correspondiente diferencial. Y como, en un plano, el incremento total es la suma de los incrementos parciales, vemos cómo el sustituir la superficie por su plano tangente permite, en primera aproximación, descomponer el incremento total sobre la superficie como la suma de los incrementos parciales, medidos sobre las correspondientes curvas coordenadas.

Obsérvese que la ecuación (3.28) tiene sentido (y define, por tanto, un plano en \mathbb{R}^3) para una función que simplemente posea derivadas parciales en (x_0, y_0) , pero, si la función no es diferenciable, el plano de ecuación (3.28) no tiene ninguna utilidad como aproximación local de la superficie. Por ejemplo, si $f(x, y)$ vale cero sobre los ejes x e y y $f(x, y) = 1$ en los demás puntos del plano, entonces $f'_x(0, 0) = f'_y(0, 0) = 0$ y la ecuación (3.28) es $Z = 0$; la diferencia $f(x, y) - Z$ vale 1 para todos los puntos del plano que no están en los ejes y ni siquiera converge a 0 cuando $(x, y) \rightarrow (0, 0)$.

Concepto de gradiente

Todavía dentro de las consideraciones geométricas, es muy interesante interpretar el término de primer orden

$$\frac{\partial f}{\partial x}(x_0, y_0)h + \frac{\partial f}{\partial y}(x_0, y_0)k$$

como el *producto escalar* del vector incremento (h, k) y el vector

$$\left(\frac{\partial f}{\partial x}(x_0, y_0), \frac{\partial f}{\partial y}(x_0, y_0) \right)$$

Este vector de \mathbb{R}^2 se llama *gradiente* de la función y se denota por $\nabla f(x_0, y_0)$ ó $\text{grad } f(x_0, y_0)$. Si f tiene derivadas parciales en todo $D \subset \mathbb{R}^2$, se llama *campo gradiente* a la aplicación de D en \mathbb{R}^2 dada por

$$(x_0, y_0) \in D \mapsto \nabla f(x_0, y_0) = \left(\frac{\partial f}{\partial x}(x_0, y_0), \frac{\partial f}{\partial y}(x_0, y_0) \right)$$

Si f es diferenciable en (x_0, y_0) , entonces la *diferencial* $df(x_0, y_0; h, k)$ es el *producto escalar del gradiente de f en (x_0, y_0) por el vector de incrementos (h, k)* :

$$df(x_0, y_0; h, k) = \nabla f(x_0, y_0) \cdot (h, k)$$

Por ejemplo, si $f(x, y) = x^2 - 2xy^3$ y $(x_0, y_0) = (1, 2)$, entonces

$$\nabla f(x, y) = (2x - 2y^3, -6xy^2) \Rightarrow \nabla f(1, 2) = (-14, -24),$$

y, para un vector de incrementos (h, k) arbitrario,

$$df(1, 2; h, k) = (-14, -24) \cdot (h, k) = -14h - 24k$$

Esta expresión de la diferencial como producto escalar tiene una aplicación interesante: si al pasar de (x_0, y_0) a $(x_0 + h, y_0 + k)$ la función f experimenta

un incremento $f(x_0 + h, y_0 + k) - f(x_0, y_0)$ y utilizamos para medirlo aproximadamente la diferencial, cabe preguntarse (y ello es útil en determinadas circunstancias) por la dirección (h, k) de \mathbb{R}^2 para la cual, dado un tamaño fijo $t = \sqrt{h^2 + k^2}$ del vector de incrementos, se obtendría un *crecimiento máximo* de f ; ello ocurrirá cuando la cantidad

$$\nabla f(x_0, y_0) \cdot (h, k) = |\nabla f(x_0, y_0)| \cdot |(h, k)| \cos \theta = t |\nabla f(x_0, y_0)| \cos \theta$$

sea máxima, siendo θ el ángulo que forman el gradiente (supuesto no nulo) y el vector de incrementos (h, k) . Como el gradiente está dado una vez fijado el punto (x_0, y_0) y t también se supone fijado, el valor máximo se alcanzará cuando $\theta = 0$ ($\cos \theta = 1$) y el valor correspondiente será $t |\nabla f(x_0, y_0)|$. Tenemos, pues, la misma propiedad que vimos en el capítulo 1 para las funciones lineales: el gradiente de f en (x_0, y_0) , si no es nulo, marca la **dirección de máximo crecimiento** de f a partir de (x_0, y_0) , si bien ahora hay que entender esto, como decimos, en *primera aproximación*, o sea, usando la diferencial como aproximación del incremento.

En el ejemplo precedente, la dirección de máximo crecimiento sería la dirección marcada por el vector $\nabla f(1, 2) = (-14, -24)$. Como $|(-14, -24)| = \sqrt{772} \simeq 27,784$, esto significa que, si aplicamos incrementos (h, k) de módulo $\sqrt{h^2 + k^2} = t$ prefijado (que podríamos interpretar como un "coste" asociado al proceso incremental), el máximo incremento que podría experimentar la función será de $27,784t$ unidades, y se lograría ese máximo incremento si (h, k) están en la dirección de $(-14, -24)$, es decir, si $h/k = (-14)/(-24)$, o, expresado en función de t , si $(h, k) = (\frac{-14t}{\sqrt{772}}, \frac{-24t}{\sqrt{772}})$.

Naturalmente, todas estas conclusiones deben entenderse siempre en términos aproximados, resultando más exactas cuanto menor sea t .

NOTA. Aquí tenemos una de las pruebas más claras de la inmensa utilidad del cálculo diferencial: sin él, la búsqueda de la dirección de máximo crecimiento supondría un trabajo enorme incluso en funciones de sólo dos variables, pues deberíamos elegir una muestra suficiente de puntos en un pequeño círculo centrado en el de partida, evaluar los incrementos correspondientes y elegir el mayor. Piénsese lo que supondría hacer este proceso para una función de muchas variables. El cálculo diferencial, sin embargo, nos resuelve el problema simplemente calculando tantas derivadas como variables independientes tenga la función.

Aplicación al cálculo aproximado

La expresión de la diferencial tiene una primera aplicación en el cálculo aproximado de incrementos, cuando no se necesita cuantificar el error cometido.

EJEMPLO. En un cierto proceso productivo se emplean x piezas de maquinaria operadas por y trabajadores, obteniéndose $f(x, y) = 30x^{1/3}y^{2/3}$ unidades

de producto. Se parte de una situación inicial de $x_0 = 1000$ máquinas e $y_0 = 8000$ trabajadores, y se desea obtener un incremento de producción del 0,2%. Analizar las siguientes posibilidades, comparando en todos los casos el incremento aproximado con el exacto.

- La empresa adquiere una nueva máquina. ¿Cuántos trabajadores adicionales debería contratar?
- Por necesidades de la producción, cada máquina nueva que se adquiera debe ser operada por dos trabajadores. ¿Cuántas máquinas nuevas habría que adquirir y cuántos trabajadores habría que contratar para producir el mismo incremento?
- ¿Cuál sería la forma más eficiente de conseguir dicho incremento?
- La empresa contrata a 40 nuevos trabajadores. ¿Cómo interpretar el resultado obtenido?

Solución: En la situación de partida se producen

$$f(x_0, y_0) = 30(1.000)^{1/3}(8.000)^{2/3} = 120.000$$

unidades de producto. El incremento buscado es del 0,2% de 120.000, es decir, de 240 unidades.

Aplicando la aproximación proporcionada por la diferencial,

$$\Delta f = f(x_0 + h, y_0 + k) - f(x_0, y_0) \simeq f_x(x_0, y_0)h + f_y(x_0, y_0)k$$

y teniendo en cuenta que

$$f_x = 30(1/3)x^{-2/3}y^{2/3} = 10(y/x)^{2/3}, \quad f_y = 30(2/3)x^{1/3}y^{-1/3} = 20(x/y)^{1/3}$$

vemos que

$$\begin{aligned} f_x(1.000, 8.000) &= 10(8.000/1.000)^{2/3} = 40 \\ f_y(1.000, 8.000) &= 20(1.000/8.000)^{1/3} = 10 \end{aligned}$$

y entonces

$$\nabla f(1.000, 8.000) = (40, 10)$$

y

$$\Delta f \simeq df(1.000, 8.000; h, k) = (40, 10) \cdot (h, k) = 40h + 10k$$

- a) En este caso, $h = 1$, luego

$$240 = 40 + 10k \Rightarrow k = 20$$

Por tanto, la empresa deberá contratar a 20 nuevos trabajadores.

- b) La segunda condición impuesta equivale a exigir que el incremento en el número de trabajadores sea el *doble* que el de máquinas, es decir: $k = 2h$. Por tanto,

$$240 = 40h + 10(2h) = 60h$$

de donde $h = 4$ (y, por tanto, $k = 8$). Es decir, habría que adquirir 4 nuevas máquinas y contratar a 8 nuevos trabajadores.

- c) La solución más eficiente desde el punto de vista de la producción (sin tener en cuenta otras consideraciones, por ejemplo, de tipo presupuestario) sería la obtenida haciendo que los incrementos se moviesen en la *dirección de máximo crecimiento*, es decir, la del *gradiente*. Como $\nabla f(x_0, y_0) = (40, 10)$, exigiremos que $h/k = 40/10$, es decir, $h = 4k$. Entonces,

$$240 = 40(4k) + 10k = 170k \Rightarrow k = \frac{240}{170} \simeq 1,4117$$

$$\text{y } h = 4k = 1,4117 \times 4 = 5,6468$$

- d) Si $k = 40$, $240 = 40h + 10 \times 40 \Rightarrow h = -4$. Esto significaría que la empresa podría prescindir de 4 máquinas y aun así conseguir el objetivo de producción propuesto.

Comparación con las soluciones exactas:

- a) Para $x = 1.001$ e $y = 8.020$,

$$f(1.001, 8.020) = 30(1.001)^{1/3}(8.020)^{2/3} \simeq 120.239,97$$

y el incremento real obtenido sería de 239,97 unidades de producto. La aproximación, pues, es francamente buena, debido al que los incrementos considerados ($h = 1, k = 20$) son muy pequeños comparados con los valores $x = 1.000, y = 8.000$ de partida.

- b) El incremento exacto de producción en este caso sería de

$$30(1.004)^{1/3}(8.008)^{2/3} - 120.000 \simeq 239,88$$

y la aproximación es ligeramente peor que en el caso anterior, aunque se mantiene dentro de unos límites absolutamente aceptables.

c) El incremento exacto sería de

$$30(1005, 6468)^{1/3}(8001, 4117)^{2/3} - 120.000 = 239, 5913.$$

d) El incremento exacto sería de $30(996)^{1/3}(8040)^{2/3} - 120000 = 238.92$.

Interpretemos geoméricamente los cálculos efectuados en este ejercicio. Recordemos que $f(x, y) = 30x^{1/3}y^{2/3}$ es una de las funciones de producción de *Cobb-Douglas*, introducidas en el primer capítulo, y cuyas curvas de nivel fueron discutidas allí. El problema puede formularse así: analizar y comparar diversas formas de pasar de la curva de nivel $f(x, y) = 120.000$, que es la que contiene al punto $(1.000, 8.000)$, a la curva de nivel $f(x, y) = 120.000 + 240 = 120.240$. Si tuviéramos un dibujo perfecto de éstas, el análisis gráfico sería, sin duda, el más eficaz para resolver el ejercicio. (Figura 3.7). En la práctica, sin embargo, no se suele disponer de mapas de curvas de nivel lo suficientemente precisos como para aplicar el método gráfico.

¿Qué sucede con las curvas de nivel cuando utilizamos la diferencial como aproximación del incremento? Como sabemos, emplear este método equivale a *sustituir la función por su aproximación lineal* en el punto dado, o, en términos geométricos, *sustituir la superficie $z = f(x, y)$ por su plano tangente $z = z_0 + f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0)$* . En este caso, *sustituir*

$$f(x, y) = 30x^{1/3}y^{2/3}$$

por

$$l(x, y) = 120.000 + 40(x - 1.000) + 10(y - 8.000) = 40x + 10y$$

(Nota: El hecho de que $l(x, y)$ carezca de término independiente, es decir, que el plano tangente pase por el origen, es meramente casual.).

Así, el uso de la diferencial equivale a *sustituir las curvas de nivel de f por las rectas de nivel de l* , sobre la base de que éstas existan, es decir, de que l no sea constante. En otras palabras, equivale a *rectificar* las curvas de nivel como si aplicásemos una "lente" al mapa de curvas de nivel de f y observásemos en su lugar un haz de rectas paralelas, las rectas de nivel de l (Figura 3.7). En el próximo capítulo demostraremos rigurosamente la existencia de dicha "lente"; de momento, limitémonos a aplicar esta idea a nuestro problema, y observar que los cálculos que hemos efectuado en el ejercicio corresponden a *resolver exactamente el problema aproximado* de pasar de la recta de nivel $l(x, y) = 120.000$ a la recta de nivel $l(x, y) = 120.240$, proceso realmente simple por tratarse de rectas, y además, *paralelas*. Aún no estamos en condiciones de estimar el error cometido en esta aproximación, y por eso lo hemos calculado



Figura 3.7

numéricamente, llegando a la conclusión de que es muy pequeño: esto sugiere que, cerca del punto $(1.000, 8.000)$, las curvas de nivel de f se aproximan bastante al haz de rectas paralelas $10x + 40y = \text{constante}$.

Terminemos esta discusión con un último comentario acerca del apartado (c) del ejercicio. Vimos en el capítulo 1 que el gradiente $\nabla l = (40, 10)$ de la función lineal $l = 40x + 10y$ representa la dirección de máximo crecimiento. Por otra parte, ∇l coincide con ∇f en el punto $(x_0, y_0) = (1.000, 8.000)$, pues, como sabemos, los coeficientes a y b en la expresión $l(x, y) = ax + by + c$ son precisamente las derivadas parciales de f en (x_0, y_0) . Por tanto, la elección de $\nabla f(1.000, 8.000)$ como dirección de máximo crecimiento se corresponde exactamente con la elección de ∇l en el problema lineal, que es el que emplearíamos para pasar de la recta $40x + 10y = 120.000$ a la paralela $40x + 10y = 120.240$ usando el segmento más corto.

Pero no acaban aquí las analogías. Como también se vio en el capítulo 1, el gradiente ∇l es perpendicular a las rectas de nivel de l . La pregunta lógica es, pues: ¿qué relación hay entre el gradiente de f y la curva de nivel de f ? En este caso, la curva de nivel de f que pasa por $(1.000, 8.000)$ se puede poner en forma explícita:

$$30x^{1/3}y^{2/3} = 120.000 \Rightarrow y^{2/3} = 4.000x^{-1/3} \Rightarrow y = (4.000)^{3/2}x^{-1/2} \stackrel{\text{def}}{=} g(x)$$

y entonces podemos calcular su pendiente: $g'(x) = -(1/2)(4.000)^{3/2}x^{-3/2}$, de donde $g'(1.000) = -4$, y vemos que la pendiente de la curva de nivel $y = g(x)$ en el punto $(1.000, 8.000)$ vale -4 , lo mismo que la pendiente de la recta de nivel $40x + 10y = 120.000$. Al coincidir ambas pendientes en el punto, deducimos la importante observación de que la curva de nivel $f(x, y) = 120.000$ y la recta de nivel $l(x, y) = 120.000$ son tangentes en el punto $(1.000, 8.000)$. La conclusión es, pues, que el gradiente de f en $(1.000, 8.000)$, perpendicular a la recta de nivel de l , es también perpendicular a la curva de nivel de f que pasa por dicho punto, entendiendo esta frase en el sentido de que es perpendicular a la tangente a la curva de nivel en dicho punto (que no es otra que la correspondiente recta de nivel de l).

En el siguiente ejercicio se comprueba que estos resultados no son privativos de las funciones de Cobb-Douglas, sino mucho más generales.

EJERCICIO: Se considera una función de la forma $f(x, y) = y^m g(x)$, donde g es derivable y m es un entero impar. Sean $y_0 \neq 0$ y x_0 tal que $g(x_0) \neq 0$.

- Obtener su aproximación lineal $l(x, y)$ en (x_0, y_0) .
- Expresar la curva de nivel de f que pasa por (x_0, y_0) en forma explícita, hallar su pendiente en (x_0, y_0) y comprobar que la curva es tangente a la recta de nivel de l que pasa por (x_0, y_0) . ¿En qué medida se utilizan las restricciones $y_0 \neq 0, g(x_0) \neq 0$? ¿Puede eliminarse alguna en algún caso?
- Comprobar que $\nabla f(x_0, y_0)$ es perpendicular a la curva de nivel que pasa por (x_0, y_0) .
- Repetir estos pasos para $f(x, y) = (y + h(x))^m g(x)$, con m impar y g y h derivables. Se supone ahora que $y_0 + h(x_0) \neq 0$ y $g(x_0) \neq 0$.
- Lo mismo con $f(x, y) = e^{y+h(x)} g(x)$ y $f(x, y) = g(x) \ln(y^m + h(x))$, indicando las condiciones que hay que exigir a (x_0, y_0) .

En este ejercicio se comprueba que la relación de tangencia entre curvas y rectas de nivel no es un hecho excepcional, al menos para funciones cuyas curvas de nivel se puedan expresar fácilmente en forma explícita al objeto de calcular sus pendientes. En el siguiente capítulo justicaremos todas estas afirmaciones para una función arbitraria $f(x, y)$, pues encontraremos un método muy sencillo para calcular la pendiente de una curva de nivel *sin necesidad de expresar ésta en forma explícita*.

Funciones de n variables

La extensión de las anteriores definiciones y resultados a funciones de n variables es inmediata, pero se recomienda al lector verificar por sí mismo todas la afirmaciones que siguen:

Definición 3.7 Sea $f: D \rightarrow \mathbb{R}$, donde $D \subset \mathbb{R}^n$, y sea $x_0 = (x_1^0, x_2^0, \dots, x_n^0)$ un punto interior de D .

- El **gradiente** de f en x_0 es el vector de \mathbb{R}^n

$$\nabla f(x_0) = \text{grad } f(x_0) = \left(\frac{\partial f}{\partial x_1}(x_0), \frac{\partial f}{\partial x_2}(x_0), \dots, \frac{\partial f}{\partial x_n}(x_0) \right) \quad (3.29)$$

2. Se dice que f es **diferenciable** en x_0 si existen las derivadas parciales $D_i f(x_0)$ y para todo vector $h = (h_1, h_2, \dots, h_n)$ de norma $|h|$ suficientemente pequeña se tiene:

$$\begin{aligned} & f(x_1^0 + h_1, x_2^0 + h_2, \dots, x_n^0 + h_n) - f(x_1^0, x_2^0, \dots, x_n^0) \\ &= \frac{\partial f}{\partial x_1}(x_0)h_1 + \dots + \frac{\partial f}{\partial x_n}(x_0)h_n + g(h_1, h_2, \dots, h_n) \end{aligned} \quad (3.30)$$

donde $g(h_1, h_2, \dots, h_n)$ es una expresión que satisface

$$\lim_{(h_1, h_2, \dots, h_n) \rightarrow (0, 0, \dots, 0)} \frac{g(h_1, h_2, \dots, h_n)}{\sqrt{h_1^2 + h_2^2 + \dots + h_n^2}} = 0 \quad (3.31)$$

es decir,

$$f(x_0 + h) - f(x_0) = \nabla f(x_0) \cdot h + g(h), \quad \text{donde} \quad \lim_{h \rightarrow 0} \frac{g(h)}{|h|} = 0. \quad (3.32)$$

Si f es diferenciable en x_0 , su **diferencial** en x_0 es la **aplicación lineal**

$$df(x_0) : (h_1, h_2, \dots, h_n) \mapsto \frac{\partial f}{\partial x_1}(x_0)h_1 + \dots + \frac{\partial f}{\partial x_n}(x_0)h_n = \nabla f(x_0) \cdot h$$

representándose sus valores por $df(x_1^0, x_2^0, \dots, x_n^0; h_1, h_2, \dots, h_n)$ ó $df(x_0; h)$.

3. El **hiperplano tangente** en x_0 a la **gráfica** de la función $f(x)$ es el **hiperplano** de \mathbb{R}^{n+1} cuya ecuación cartesiana es

$$z = f(x_1^0, x_2^0, \dots, x_n^0) + \frac{\partial f}{\partial x_1}(x_0)(x_1 - x_1^0) + \dots + \frac{\partial f}{\partial x_n}(x_0)(x_n - x_n^0)$$

es decir,

$$z = f(x_0) + \nabla f(x_0) \cdot (x - x_0)$$

llamando $(x_1, x_2, \dots, x_n, z)$ a las componentes de los vectores de \mathbb{R}^{n+1} .

Una aplicación lineal de \mathbb{R}^n en \mathbb{R} tiene la forma

$$L(h) = a_1 h_1 + a_2 h_2 + \dots + a_n h_n = \langle a, h \rangle \quad (3.33)$$

para un cierto vector $a = (a_1, a_2, \dots, a_n) \in \mathbb{R}^n$ unívocamente determinado. Como para las funciones de dos variables, se prueba en el caso general que si f es diferenciable en x_0 , la diferencial

$$df(x_0) : \mathbb{R}^n \ni h \mapsto \nabla f(x_0) \cdot h = \frac{\partial f}{\partial x_1}(x_0)h_1 + \dots + \frac{\partial f}{\partial x_n}(x_0)h_n \in \mathbb{R}$$

es la única aplicación lineal de \mathbb{R}^n en \mathbb{R} que aproxima al incremento $f(x_0+h) - f(x_0)$ salvo "infinitésimos de orden superior", o sea, que satisface la condición (3.31), la cual se expresa en forma concisa por " $f(x_0+h) - f(x_0) - \nabla f(x_0) \cdot h = o(|h|)$ cuando $|h| \rightarrow 0$ ".

Que la función f es diferenciable en x_0 , por tanto, equivale a decir que existe una única aplicación lineal $L: \mathbb{R}^n \rightarrow \mathbb{R}$, dependiente, en general, de x_0 y definida por una expresión de la forma (3.33) tal que

$$\lim_{h \rightarrow 0} \frac{f(x_0+h) - f(x_0) - L(h)}{|h|} = 0 \quad (3.34)$$

Se puede adoptar este enunciado como definición de función diferenciable y, a partir de él, se demuestra que f tiene derivadas parciales en x_0 y que L tiene la forma

$$L(h) = \frac{\partial f}{\partial x_1}(x_0)h_1 + \dots + \frac{\partial f}{\partial x_n}(x_0)h_n$$

También podemos formular la definición de diferenciable con la expresión alternativa del **resto** g :

Definición 3.8 f es diferenciable en x_0 si el incremento $f(x_0+h) - f(x_0)$ se puede expresar en la forma

$$f(x_0+h) - f(x_0) = \nabla f(x_0) \cdot h + h_1\delta_1(h) + \dots + h_n\delta_n(h) \quad (3.35)$$

donde $\delta_i(h)$ son funciones tales que $\lim_{h \rightarrow 0} \delta_i(h) = 0$.

Utilizando (3.35) podemos demostrar, como antes, que toda función diferenciable en un punto es continua en él. Pero veamos la continuidad de f en x_0 por una vía alternativa que nos proporciona una estimación del incremento de la función. Como para el caso de dos variables, la propiedad (3.31) de la función **resto** g se puede expresar en la forma equivalente: dado $\varepsilon > 0$, existe $\delta > 0$ de modo que

$$|g(h)| \leq \varepsilon |h| \text{ para todo } h \text{ tal que } |h| \leq \delta \quad (3.36)$$

(Obsérvese que (3.36) se verifica obviamente para $h = 0 = (0, \dots, 0)$ pues $g(0) = 0$ por (3.30)). Tomando $\varepsilon = 1$ en (3.36), existe $\delta > 0$ tal que, si $|h| \leq \delta$,

$$\begin{aligned} |f(x_0+h) - f(x_0)| &\leq |\nabla f(x_0)| |h| + |g(h)| \leq |\nabla f(x_0)| |h| + 1 \cdot |h| = \\ &= (|\nabla f(x_0)| + 1) |h| \end{aligned}$$

desigualdad de la que se deriva inmediatamente la continuidad en x_0 .

Nos queda pendiente la demostración de la condición suficiente de diferenciable:

Teorema 3.2 Si f tiene derivadas parciales continuas en un conjunto abierto $D \subset \mathbb{R}^n$, entonces es diferenciable en D .

Demostración: Hay que probar para cada $x_0 \in D$ que

$$\lim_{|h| \rightarrow 0} \frac{f(x_0 + h) - f(x_0) - \nabla f(x_0) \cdot h}{|h|} = 0 \quad (3.37)$$

La idea de la demostración consiste en descomponer el incremento $f(x_0 + h) - f(x_0)$ en una suma "telescópica" de modo que en cada sumando sólo varíe una coordenada y utilizar sucesivamente, para cada uno de ellos, el teorema del valor medio para funciones de una variable: dado $h = (h_1, h_2, \dots, h_n)$ definimos los vectores

$$h^{(k)} = \sum_{i=1}^k h_i e_i, \quad k = 1, \dots, n, \quad h^{(0)} = (0, \dots, 0), \quad h^{(n)} = h \quad (3.38)$$

(o sea, $h^{(1)} = (h_1, 0, \dots, 0)$, $h^{(2)} = (h_1, h_2, 0, \dots, 0)$, etcétera; e_i son los vectores de la base canónica de \mathbb{R}^n ; véase la ilustración de la figura 3.8 para $n = 2$). Se tiene:

$$f(x_0 + h) - f(x_0) = \sum_{k=1}^n [f(x_0 + h^{(k)}) - f(x_0 + h^{(k-1)})] \quad (3.39)$$

Fijado $x_0 \in D$, existe $r > 0$ tal que $B_r(x_0) \subset D$. Si $|h| < r$, se tendrá también $|h^{(k)}| < r$, ya que $|h^{(k)}| \leq |h|$; en consecuencia, $x_0 + h^{(k)} \in B_r(x_0) \subset D$ para todo k y, por ser la bola $B_r(x_0)$ un conjunto convexo, el segmento $[x_0 + h^{(k-1)}, x_0 + h^{(k)}]$ está completamente contenido en $B_r(x_0) \subset D$. Por el teorema del valor medio aplicado a cada función parcial $x_k \mapsto f(x_1^0, \dots, x_k, \dots, x_n^0)$, existe un punto $\xi^{(k)} \in D$ perteneciente al segmento que une $x_0 + h^{(k-1)}$ y $x_0 + h^{(k)}$ de modo que

$$f(x_0 + h^{(k)}) - f(x_0 + h^{(k-1)}) = h_k \frac{\partial f}{\partial x_k}(\xi^{(k)}) \quad (3.40)$$

Así pues,

$$f(x_0 + h) - f(x_0) = \sum_{k=1}^n h_k \frac{\partial f}{\partial x_k}(\xi^{(k)}) \quad (3.41)$$

de donde

$$f(x_0 + h) - f(x_0) - \nabla f(x_0) \cdot h = \sum_{k=1}^n h_k \left[\frac{\partial f}{\partial x_k}(\xi^{(k)}) - \frac{\partial f}{\partial x_k}(x_0) \right]$$

y

$$\begin{aligned} \frac{|f(x_0 + h) - f(x_0) - \nabla f(x_0) \cdot h|}{|h|} &\leq \sum_{k=1}^n \frac{|h_k|}{|h|} \left| \frac{\partial f}{\partial x_k}(\xi^{(k)}) - \frac{\partial f}{\partial x_k}(x_0) \right| \leq \\ &\leq \sum_{k=1}^n \left| \frac{\partial f}{\partial x_k}(\xi^{(k)}) - \frac{\partial f}{\partial x_k}(x_0) \right| \end{aligned}$$

Cuando $h \rightarrow 0$, los $\xi^{(k)}$ tienden a x_0 , y, entonces, por la continuidad de las derivadas parciales,

$$\frac{\partial f}{\partial x_k}(\xi^{(k)}) - \frac{\partial f}{\partial x_k}(x_0) \rightarrow 0$$

lo que implica (3.37).

(Recuérdese que en el teorema del valor medio para funciones de una variable no se exige que la derivada de la función sea continua; así que, de la fórmula (3.41), se deduce un resultado interesante: Si f tiene derivadas parciales en un conjunto abierto $D \subset \mathbb{R}^n$ y son funciones acotadas, entonces f es continua en D ; en realidad, no se necesita la acotación global en D ; basta suponer que las derivadas parciales sean localmente acotadas, o sea, que para cada punto de D exista un ε -entorno en el cual estén acotadas. Por otra parte, se puede demostrar que si f tiene derivadas parciales y $n-1$ de ellas son continuas en D , entonces f es diferenciable en D).

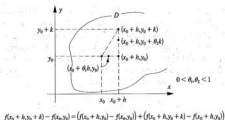


Figura 3.8

Definición 3.9 Sea $D \subset \mathbb{R}^n$ abierto. Una función $f : D \rightarrow \mathbb{R}$ se dice que es **diferenciable con continuidad**, o **continuamente diferenciable**, en D si existen las derivadas parciales $\partial f / \partial x_i$ y son funciones continuas en todo D . El conjunto de tales funciones se denota por $C^1(D)$ y, en consonancia con ello, se las conoce también como **funciones de clase uno** (o de clase C^1) en D .

3.4 Notas al capítulo

1. Nota importante sobre la notación.

Si la elección de una buena notación es importante en todos los ámbitos de las matemáticas, en el caso del cálculo diferencial es todavía más fundamental, dada la complejidad y nivel de abstracción de los conceptos que maneja. Sin embargo, así como en otras áreas matemáticas, como el álgebra o la geometría, se ha ido imponiendo a lo largo del tiempo un sistema establecido y comúnmente aceptado de notación, en el cálculo diferencial se da la peculiaridad de que las notaciones modernas no han llegado a desplazar a las antiguas, con la consecuencia de que, actualmente, coexisten diversas notaciones totalmente distintas entre sí, tanto en su aspecto externo como en su espíritu, y la razón de tal diversidad es que cada una de ellas tiene sus ventajas y sus inconvenientes. Los libros modernos de Cálculo Diferencial dirigidos a matemáticos utilizan exclusivamente las llamadas *notaciones modernas*, lo cual priva a sus lectores el acceso a muchos libros o trabajos procedentes de las ciencias aplicadas —que en su mayoría siguen empleando las notaciones tradicionales—, así como también dificulta su lectura el uso de la terminología —y metodología— de las *magnitudes infinitesimales*, desaparecida desde hace mucho tiempo de los textos de matemáticas por y para matemáticos. Por este motivo trataremos aquí de combinar los diversos sistemas de notación, con el objeto de que el lector se familiarice con todos ellos y entienda los pros y contras de cada sistema: las notaciones modernas aportan precisión y exactitud, aun a costa de recargarlas hasta extremos ilegibles en muchas ocasiones; mientras que el sistema tradicional es más breve y conciso, pero frecuentemente cae en ambigüedades y abusos de notación tan excesivos que, en ocasiones, no es posible descifrar el sentido de sus fórmulas y expresiones.

Una de las diferencias más acusadas es la siguiente: la notación tradicional pone el énfasis en las *variables*, omitiendo muchas veces toda referencia a la propia función que las liga, dándola por sobreentendida. Así, la derivada de $y = f(x)$ se representa simplemente por y' ó dy/dx , entendiéndose que el contexto aclara de qué función se trata. No es que se pretenda indicar que se está derivando una variable —cosa sin sentido hoy día, aunque en los orígenes del cálculo se pensase en esos términos— sino que se está abreviando la notación más de lo debido. Ni que decir tiene que si tenemos las mismas dos variables o magnitudes ligadas por diversas relaciones $y = f(x)$, $y = g(x)$, etc., la confusión puede ser total. Una buena solución es la introducida por Lagrange, $f'(x)$, $g'(x)$, ..., que evita este problema obligando a incorporar el símbolo (f ó g) de la función de que se trata. La notación moderna $Df(x_0)$ es equivalente a ésta a todos los respectos. Sin embargo, dado que en las ciencias aplicadas lo más natural es manejar “variables” como representativas de “magnitudes” con

contenido físico, económico, etc., dándose por sobreentendidas las relaciones funcionales existentes entre dichas magnitudes; la tradición de representar, por ejemplo, la derivada de la renta Y con respecto al tiempo t como dY/dt ó Y' (¡omitendo incluso la variable independiente!) está firmemente establecida, y parece difícil de cambiar.

Otra fuente frecuente de confusión es la omisión de toda referencia a los valores en los que se evalúan las funciones. Así, en una relación funcional $y = f(x)$, se sobreentiende que y_0 es el valor correspondiente a un cierto x_0 , que Δy ó Δf es el incremento de y ó de f correspondiente a un Δx que nunca ha sido mencionado antes, que cuando se escribe y' se está evaluando la derivada en un cierto x "genérico" que se supone conocido o reconocible por el contexto, etc.

La pregunta inmediata es: ¿por qué mantener estos convenios notacionales, a sabiendas de que pueden (¡y suelen!) dar lugar a confusión? Bien, demos simplemente un ejemplo: la *regla de la cadena* o de derivación de funciones compuestas que se estudia en el curso de funciones de una variable. En notación moderna, dadas dos funciones de una variable f y g , la derivada de la función compuesta $x \mapsto f[g(x)]$, que se denota $f \circ g$, viene dada por la fórmula $(f \circ g)'(x_0) = f'[g(x_0)] \cdot g'(x_0)$ -o bien: $D(f \circ g)(x_0) = (Df)[g(x_0)] \cdot Dg(x_0)$ -. Si introducimos nombres para las variables y llamamos $y = g(x)$, $z = f(y)$, la misma fórmula en la notación tradicional de LEIBNIZ (a quien, por cierto, se debe la fórmula) se lee

$$\frac{dz}{dx} = \frac{dz}{dy} \cdot \frac{dy}{dx}$$

¡que incluso parece evidente!

En resumen, hay razones de peso que han venido bloqueando la utilización sistemática de las notaciones modernas en las ciencias aplicadas. Por todo ello, creemos que es de interés para el lector interesado en éstas el saber interpretar bien las (a menudo crípticas) notaciones tradicionales, pues son las que, para bien o para mal, se va a encontrar en la mayor parte de la bibliografía especializada de su ámbito de estudio.

2. La "expresión analítica" de la diferencial.

Como continuación e ilustración de la nota anterior, vamos a presentar un tema obligado en el desarrollo clásico del cálculo diferencial y que constituye una de las consecuencias más llamativas del intercambio indiscriminado de papeles entre funciones y variables tan característico de las notaciones y métodos tradicionales. Recordemos la definición de diferencial de una función de dos variables $f: \mathbb{R}^2 \rightarrow \mathbb{R}$

$$df(x_0, y_0; h, k) = \frac{\partial f}{\partial x}(x_0, y_0)h + \frac{\partial f}{\partial y}(x_0, y_0)k$$

y eliminemos toda referencia a los valores en los que se evalúan las funciones, llamando Δx a h , Δy a k para mejor recordar su papel en la fórmula anterior:

$$df = \frac{\partial f}{\partial x} h + \frac{\partial f}{\partial y} k \quad \text{o} \quad df = \frac{\partial f}{\partial x} \Delta x + \frac{\partial f}{\partial y} \Delta y$$

Eliminemos ahora de la fórmula a la propia función f , llamando z a $f(x, y)$:

$$dz = \frac{\partial z}{\partial x} h + \frac{\partial z}{\partial y} k \quad \text{o} \quad dz = \frac{\partial z}{\partial x} \Delta x + \frac{\partial z}{\partial y} \Delta y$$

obteniendo así la forma usual de definir la diferencial en notación tradicional. Hasta ahora el proceso es claro y el paso de una a otra notación es bastante sencillo (¡y reversible!). Ahora, de la mano algo críptica de Leibniz, aplicamos esta fórmula a la función $z = x, y$, dado que $\partial z / \partial x = 1$, $\partial z / \partial y = 0$, vemos que $dx = 1 \cdot \Delta x + 0 \cdot \Delta y = \Delta x$, y si hacemos $z = y$, obtenemos $dy = \Delta y$. Estas fórmulas se explican tradicionalmente con frases del tipo: "... por lo tanto, las diferenciales de las variables independientes coinciden con sus propios incrementos..." , aclaración que, por supuesto, no es tal, pues las variables independientes no tienen derivadas, ni menos diferenciales: recordemos que, en la teoría que venimos desarrollando, los únicos objetos matemáticos susceptibles de derivación son las *funciones* o aplicaciones de \mathbb{R} ó \mathbb{R}^n en \mathbb{R} , y no las variables, que no son otra cosa que letras que representan elementos genéricos de \mathbb{R} ó de \mathbb{R}^n . Pero no termina ahí la cosa, pues la idea clave de todo este proceso es la inversa: sustituir las fórmulas $dx = \Delta x$, $dy = \Delta y$ en la definición de diferencial, obteniendo

$$dz = \frac{\partial z}{\partial x} dx + \frac{\partial z}{\partial y} dy \quad (3.42)$$

fórmula conocida como **"expresión analítica de la diferencial"**. En el cálculo de una variable, si $z = f(x)$, la definición $dz = z' \Delta x$ particularizada para la función $z = x$ daría $dx = \Delta x$, con lo cual tendríamos $dz = z' dx$, de donde la famosa notación dz/dx para z' , y la no menos famosa explicación "... y, por tanto, la derivada [de una función de una variable] es un verdadero cociente de diferenciales..." En dos variables, ya se ve que dz/dx no coincide con $\partial f / \partial x$, sino con $\partial f / \partial x + (\partial f / \partial y)(dy/dx)$ (expresión que admitirá interpretación en el próximo capítulo ¡no se lo pierda!), por lo que las derivadas parciales *no* son cocientes de diferenciales: por eso se introduce la híbrida notación $\partial z / \partial x$, etc (debida a JACOBI), en la que ∂x y ∂z no tienen sentido independiente uno del otro, y la escritura en forma de cociente no tiene otro objeto que asemejarse lo más posible al caso de una variable.

¿Cómo se escribiría (3.42) en notación moderna? Lo primero que habría que hacer es dar nombres a las *funciones* cuyas diferenciales aparecen en ella.

Ya tenemos f para la relación entre z y (x, y) . Llamemos X a la función $X(x, y) = x$, e Y a la función $Y(x, y) = y$. Estas son funciones lineales cuyas diferenciales son, simplemente,

$$dX(x_0, y_0; h, k) = h, \quad dY(x_0, y_0; h, k) = k$$

o, más brevemente,

$$dX = h = \Delta x, \quad dY = k = \Delta y$$

La expresión analítica de la diferencial queda, pues, así:

$$df(x_0, y_0; h, k) = \frac{\partial f}{\partial x}(x_0, y_0) dX(x_0, y_0; h, k) + \frac{\partial f}{\partial y}(x_0, y_0) dY(x_0, y_0; h, k) \quad (3.43)$$

o bien

$$df = \frac{\partial f}{\partial x} dX + \frac{\partial f}{\partial y} dY$$

Así pues, es posible escribir la expresión analítica de la diferencial en las notaciones modernas (o semimodernas, como (3.43)). Pero, ¿qué ganamos con todo este desarrollo teórico? ¿Para qué puede servir una expresión como ésta? En la notación moderna, no parece más que una trivialidad, pues viene a decir que una función como $l(x, y) \stackrel{\text{def}}{=} Ah + Bk$ se puede escribir como combinación lineal de las funciones $X(h, k) \stackrel{\text{def}}{=} h$ e $Y(h, k) \stackrel{\text{def}}{=} k$, de la forma obvia $l = AX + BY$. La verdadera naturaleza de la fórmula se aprecia mucho mejor volviendo a la notación tradicional, y planteándose, con Leibniz, la siguiente pregunta: ¿Qué sucedería si x e y no fuesen variables independientes, sino funciones de una cuarta variable, digamos u ? Por ejemplo, si $z = x^2 + xy$, tendríamos

$$\begin{aligned} dz &= (\text{definición}) = (2x + y)\Delta x + x\Delta y = \\ &= (\text{identificación } \Delta x = dx, \Delta y = dy) = (2x + y)dx + xdy \end{aligned}$$

Supongamos ahora que $x = u^2$, $y = 3u$. Entonces tendríamos $dx = x'_u du = 2u du$, $dy = y'_u du = 3 du$ y sustituyendo en la expresión de dz nos daría

$$\begin{aligned} dz &= (2x + y)(2u du) + x(3 du) = (2(2x + y)u + 3x) du = \\ &= (2(2u^2 + 3u)u + 3u^2) du = (4u^3 + 9u^2) du, \end{aligned}$$

expresión que, de forma un tanto sorprendente, coincide exactamente con la que se obtendría sustituyendo directamente x e y en la definición de z

($z = (u^2)^2 + (u^2)(3u) = u^4 + 3u^3$) y calculando la diferencial de z como función de u : $dz = z'_u du = (4u^3 + 9u^2)du$. Este ejemplo ilustra un resultado general que demostraremos en el próximo capítulo, la llamada "*propiedad de invariancia de la diferencial*", según la cual su expresión analítica es válida tanto si dx, dy representan incrementos $\Delta x, \Delta y$ de variables independientes como si representan diferenciales de funciones $x = \phi(u, v, \dots)$, $y = \psi(u, v, \dots)$ de otras variables. Ésta constituye, probablemente, el mérito más importante del sistema de notación de Leibniz, e ilustra perfectamente el por qué de esa enorme resistencia a abandonar, en aras de la precisión en los razonamientos, una notación y una metodología que no cesan de causar asombro por su capacidad de producir fórmulas correctas basándose en un esquema teórico lleno de conceptos más que confusos.

Manteniéndonos en nuestra idea de dar al lector los medios para interpretar correctamente cualquier tipo de fórmulas del cálculo diferencial, seguiremos presentando la teoría en los términos más precisos posibles ("modernos") en el bloque principal del texto, y reservaremos para notas como ésta los comentarios sobre el sistema tradicional.

3. Nota histórica sobre el cálculo de incrementos infinitesimales.

Como ya comentamos en la Introducción al capítulo, las interpretaciones geométrica (Descartes, Pascal, Fermat, Leibniz) y dinámica (Newton) de la derivada son fundamentales para entender el tipo de argumentos que se utilizan en el cálculo diferencial. Sin embargo, no haríamos justicia a los creadores del Cálculo si nos limitásemos a estas interpretaciones sin intentar explicar qué es lo que realmente tenían *in mente* estos autores.

El objetivo fundamental es, por encima de todo, *calcular incrementos*, es decir, estudiar Δy como función de $h = \Delta x$. Por ejemplo, si $y = x^3$, $\Delta y = (x+h)^3 - x^3 = 3x^2h + 3xh^2 + h^3$, y si $y = x^n$, la fórmula del binomio de Newton nos daría una expresión polinómica para Δy que incluiría términos en h^n inclusive. Y he aquí la idea básica: sin tomar al pie de la letra lo que vamos a decir, es útil visualizar esta expresión polinómica en potencias de h como si fuese la *descomposición decimal* de un número real. Por ejemplo, $\pi = 3,141592\dots = 3 + 1 \times (0,1) + 4 \times (0,1)^2 + 1 \times (0,1)^3 + \dots$. La diferencia fundamental entre ambas expresiones es que, a diferencia del desarrollo decimal, al cálculo diferencial no le interesa fijar $h = 0,1$, sino todo lo contrario: se trata de *estudiar la dependencia de Δy como función de $h = \Delta x$* , sabiendo que:

- Al menos en estos ejemplos sencillos, la relación funcional entre ambos es de tipo polinómico.
- Interesa, sobre todo, describir dicha relación para valores pequeños de h .

Volviendo al ejemplo del área del cuadrado ($y = x^2$), si pudiéramos afirmar sin más que, Δy es igual a $2xh$, eliminando totalmente el término h^2 , tendríamos una descripción mucho más simple de la dependencia de Δy en función de h : si, por ejemplo, tuviésemos que hallar el incremento de lado h que produce un incremento deseado de área Δy (digamos $\Delta y = 0,01$) para un cuadrado de lado $x = 1$, es mucho más sencillo resolver la ecuación $2h = 0,1$ que la que realmente deberíamos resolver, a saber, $2h + h^2 = 0,1$. Dejamos al cuidado del lector comparar ambas soluciones y decidir cuántos decimales exactos de la verdadera solución se consiguen resolviendo la ecuación aproximada $2h = 0,1$, y le estimulamos a que experimente con otros valores de Δf , como 0,01, 0,001, etc. El lector más ambicioso podrá analizar el mismo problema en el caso del volumen del cubo, es decir, con la función $y = x^3$, reteniendo en la fórmula de Δy sólo el término de primer grado en h , es decir, $3x^2h$; necesitará, eso sí, consultar en algún texto de álgebra la fórmula de la solución de las ecuaciones de tercer grado...

Es obvio que la eliminación sin más del término h^2 , o de los términos $3xh^2 + h^3$, es ilegítima, salvo en el caso trivial $h = 0$. Y aun así, en los orígenes del cálculo diferencial se discutió ampliamente la posibilidad de dar validez a dicha eliminación a condición de que h fuese *infinitesimal*, noción nunca precisada claramente (y con razón: ¿qué es un número real infinitamente pequeño? ¿Cuáles son sus cifras decimales?). También se pensó en la posibilidad de extender el campo de los números reales añadiendo ciertas cantidades no reales, los *infinitésimos* (y sus recíprocos, los *infinitos*) con los cuales las mencionadas operaciones de eliminación fuesen válidas. Al fin y al cabo, la construcción de los números complejos, que es más o menos contemporánea a la invención del cálculo diferencial, no consiste en otra cosa que en ampliar el campo real con ciertos objetos "imaginarios". Y, por ejemplo, EULER nunca definió el número e mediante un límite, sino directamente como $(1 + 1/N)^N$, donde N es *infinito* [palabras textuales]. Pues bien, a pesar de siglos de esfuerzos en esta dirección, nunca se llegó a construir un sistema numérico ampliado en este sentido que no naufragase en continuas contradicciones, y la solución finalmente dada a este problema (los "números hiperreales" introducidos por ABRAHAM ROBINSON en 1960) es de tal complejidad teórica que queda totalmente fuera del alcance de un libro de estas características.

En resumen, la clave está en considerar h como una cantidad *variable*, y no como una "constante infinitesimal" como se había venido haciendo hasta entonces. Con esta idea, el punto de vista actual consiste en reconocer que la omisión del término h^2 de la fórmula de Δf no puede dar lugar a resultados correctos, aunque sí *aproximados* si h es pequeño, y tanto más aproximados cuanto más pequeño sea h . La evaluación del margen de variabilidad admisible para h y la validez de las conclusiones extraídas a partir del valor aproximado

de Δf son, precisamente, los temas esenciales del cálculo diferencial, particularmente el segundo: ¿cuándo y en qué contexto podemos garantizar que la aproximación $\Delta f = 2xh$ es “suficientemente buena”? Indiquemos finalmente que el resultado más asombroso de toda esta empresa es la posibilidad de utilizar y combinar los valores calculados con estos métodos asumiéndolos aproximados y *obtener resultados exactos* (por ejemplo, en la teoría de la optimización o en el cálculo de áreas de figuras curvilíneas, mediante el cálculo integral).

Para ver cómo se puede llevar a cabo este proyecto de “estudiar incrementos”, recordemos que, en general, si $f(x)$ es un polinomio, Δf es un polinomio en h del mismo grado cuyo término independiente es nulo, es decir:

$$\Delta f = A(x)h + B(x)h^2 + C(x)h^3 + \dots$$

(donde hemos escrito $A(x), B(x), C(x), \dots$ en vez de A, B, C, \dots para dejar clara su dependencia de x). El problema es calcular estos coeficientes. La ausencia de término independiente permite dividir por h sin salirnos del ámbito polinómico:

$$\frac{\Delta f}{h} = A(x) + B(x)h + C(x)h^2 + \dots$$

y, una vez más, el coeficiente A o $A(x)$ (llamado *primer coeficiente diferencial*) se obtiene simplemente eliminando h (o sea, haciendo $h = 0$) en la fórmula precedente. Concluimos así que, entre las funciones polinómicas, el desarrollo del incremento Δf en potencias del incremento h (que llamaremos más adelante *desarrollo de Taylor* de la función $y = f(x)$) se obtiene mediante un procedimiento puramente algebraico, basado exclusivamente en la utilización de la fórmula del binomio de Newton junto con operaciones elementales bastante evidentes, aunque laboriosas. Dado que, por ejemplo, los problemas de máximos y mínimos se pueden resolver conociendo el mencionado desarrollo de Δf , podríamos decir que, si las únicas funciones que fuésemos a manejar fuesen de tipo polinómico, no necesitaríamos introducir ninguna metodología especial ni idea nueva para abordar las cuestiones relativas a optimización.

Veamos ahora un ejemplo de función no polinómica, aunque sí racional. ¿Qué sucede para $f(x) = 1/x$? En este caso,

$$\frac{\Delta f}{h} = \frac{\frac{1}{x+h} - \frac{1}{x}}{h} = \frac{\frac{-h}{(x+h)x}}{h} = \frac{-1}{(x+h)x}$$

y podemos hacer $h = 0$ en la fórmula anterior, obteniendo $A = -1/x^2$, gracias, una vez más, a la simplificación en numerador y denominador del término h en la última igualdad. Parece, pues, posible extender el método al ámbito de las funciones racionales.

Ahora bien, ¿qué hacemos con $f(x) = \sin x$? Si simplemente hacemos $h = 0$ en el cociente incremental

$$\frac{\Delta f}{h} = \frac{\sin(x+h) - \sin x}{h}$$

todo lo que obtenemos es $0/0$. Para proceder como en el caso polinómico, deberíamos hallar un factor h en el numerador, que pudiese cancelarse con el h del denominador. Pero, ¿cómo hacerlo?

Es el concepto matemático de *límite*, analizado en el curso de funciones de una variable, el que resuelve, desde el punto de vista teórico, el problema de hallar el "verdadero valor" de la fracción anterior para $h = 0$. En otras palabras,

$$A = \lim_{h \rightarrow 0} \frac{\sin h}{h}$$

límite cuyo valor exacto es conocido en este caso ($= 1$). Es interesante que la primera vez que aparece la palabra *límite* en el sentido que le damos hoy es en un epígrafe con ese nombre, escrito por *d'Alembert* —por encargo de *Diderot*— para la famosa *Enciclopedia Francesa* de 1751. Vemos así que el avance principal hacia la fundamentación lógica del cálculo diferencial es una empresa "enciclopedista"...

Conviene saber que esta forma de llegar al concepto de derivada como *coeficiente diferencial* se corresponde con la versión dada por *LAGRANGE*, en la que la palabra *derivada* aparece por primera vez en su sentido originario de *deducida*:

"... desarrollando $f(x+h) - f(x)$ en serie

$$Ah + Bh^2 + Ch^3 + \dots$$

en la que las cantidades A, B, C, \dots , los coeficientes de las potencias de h , serán nuevas funciones de x , *derivadas* [*deducidas*] de la función *primitiva* $f(x)$ e independientes de la cantidad h .

La formación y cálculo de estas diversas funciones es, a decir verdad, el verdadero objeto del nuevo cálculo, llamado *diferencial*..."

Lagrange pretendía así eludir el problema del paso al límite. La dificultad, naturalmente, estriba en que habría que saber *a priori* que el incremento Δf se puede expresar en potencias de h , lo cual no es cierto en todos los casos, aunque sí lo es en los más importantes.

3.5 Problemas

1. Dada $f(x, y) = xy^2$, hallar $\partial f / \partial x$ en el punto $(1, 2)$ aplicando la definición de derivada parcial. Lo mismo para un punto arbitrario (x_0, y_0) .
2. Hallar $\partial f / \partial x$, $\partial f / \partial y$ (y $\partial f / \partial z$ en su caso) para las siguientes funciones:

$$\begin{array}{ll} \text{(a)} \ x^5y^3 - 3x^2y^4 + 5x^3 - 7y^2 & \text{(b)} \ \frac{x^3y}{x^2 + y^4 + 1} \\ \text{(c)} \ \ln(x^2 + y^2 + e^{xy}) & \text{(d)} \ x^4y^3z^2 \\ \text{(e)} \ \frac{1}{1 + x^2y^4z^6} & \text{(f)} \ e^{xyz} \end{array}$$

3. De una función $f(x, y)$ sólo se conocen sus valores en los puntos de la forma $(1, y)$, que resultan ser $f(1, y) = \sin y$. ¿Qué derivadas parciales de f pueden calcularse y en qué puntos? ¿Pueden obtenerse las curvas de nivel de f ?
4. Probar que la función $f(x, y) = e^{-xy}$ es decreciente en x para cada $y > 0$ fijo. ¿Cuáles son las curvas de nivel de dicha función? ¿Qué pasa si $y < 0$?
5. Para cada una de las funciones

$$\begin{array}{ll} f(x, y) = \frac{x^3y}{x^2 + y^2} & \text{si } (x, y) \neq (0, 0), \quad f(0, 0) = 0 \\ f(x, y) = xy \ln(x^2 + y^2) & \text{si } (x, y) \neq (0, 0), \quad f(0, 0) = 0 \\ f(x, y) = \frac{x^3}{x^2 + y^2} & \text{si } (x, y) \neq (0, 0), \quad f(0, 0) = 0 \end{array}$$

se pide estudiar las siguientes cuestiones:

- (a) Continuidad de f en cada punto (x, y) de \mathbb{R}^2 .
 - (b) Existencia y continuidad de las derivadas parciales en cada punto $(x, y) \in \mathbb{R}^2$.
 - (c) Diferenciabilidad de f en cada $(x, y) \in \mathbb{R}^2$.
6. Para las siguientes funciones, calcular sus derivadas parciales en un punto genérico (x, y) , y dibujar aproximadamente su vector gradiente (caso de existir) en los puntos $(0, 0)$, $(0, 1)$ y $(1, 0)$:

$$\begin{array}{ll} \text{(a)} \ x^2y^3 + \cos(xy) & \text{(b)} \ e^{x+y} + \ln(x+1) \\ \text{(c)} \ \frac{xy - x^2\sqrt{1+y}}{x^2 + y^2} & \text{(d)} \ x^ay^{1-a}, \ 0 < a < 1 \text{ constante} \\ \text{(e)} \ \frac{x}{y-1} & \text{(f)} \ x^y \end{array}$$

7. Para las siguientes funciones, dibujar aproximadamente las curvas de nivel, calcular el gradiente en el punto $(0,1)$, comprobar que éste es ortogonal a la curva de nivel que pasa por $(0,1)$ y escribir la ecuación del plano tangente en $(0,1, f(0,1))$ a la superficie $z = f(x,y)$:

- | | |
|---|-----------------|
| (a) $2x + y$ | (b) $y - x^2$ |
| (c) xy | (d) $x^2 + y^2$ |
| (e) $x^a y^{1-a}$, $0 < a < 1$ constante | |

Lo mismo para el punto $(1,2)$.

8. Dibujar el gradiente de $f(x,y,z) = x - y - z(x+y)$ en los puntos $(0,0,0)$ y $(0,1,0)$. ¿En qué puntos (x,y,z) se anula el gradiente? Lo mismo para la función $f(x,y,z) = x + y - z(x+y)$.

9. Obtener el gradiente de la función $f(x,y) = xy^2$. Dibujarlo para los puntos $(1,0)$, $(0,1)$, $(1/\sqrt{2}, 1/\sqrt{2})$. Calcular la norma $|\nabla f|$ en un punto arbitrario (x,y) . Determinar los puntos del círculo unidad $x^2 + y^2 = 1$ en los cuales dicha norma alcanza un valor máximo o mínimo. (Indicación: es equivalente obtener los máximos y mínimos del cuadrado de la norma $|\nabla f|^2$, y para ello basta sustituir y^2 por $1 - x^2$ y resolver el problema así planteado, que sólo contiene funciones de una variable).

Obtener los puntos (x,y) en los cuales el gradiente: a) es horizontal; b) es vertical; c) forma un ángulo de 45° con la horizontal; d) forma un ángulo de 30° con la horizontal.

10. Calcular aproximadamente

$$\sqrt{\sin^2(1,6) + 3e^{-0,02}}$$

utilizando el valor de la función

$$f(x,y) = \sqrt{\sin^2 x + 3e^y}$$

en el punto $(\pi/2, 0)$ ($\pi/2 \simeq 1,571$).

11. (a) Obtener la diferencial de $z = f(x,y) = x + y^2$ en el punto $(1,1)$ para los incrementos $\Delta x = 0,1$, $\Delta y = 0,2$. Calcular el incremento exacto Δz correspondiente a los valores dados. ¿Cuánto vale $\Delta z - dz$?
- (b) Lo mismo para los valores $\Delta x = 0,01$, $\Delta y = 0,02$, y, en general, para dos incrementos ligados por la relación $\Delta y = 2\Delta x$.
- (c) Si los incrementos están ligados por la relación $\Delta y = a\Delta x$, ¿para qué valores de a se puede garantizar que Δz será positivo siempre que lo sea Δx ?

12. Obtener la diferencial de $z = f(x, y) = x^2y$ en el punto $(1, 1)$ para los incrementos $\Delta x = 0,1$, $\Delta y = 0,2$. Calcular el incremento exacto Δz correspondiente a los valores dados. ¿Cuánto vale $\Delta z - dz$? ¿Y el error relativo $(\Delta z - dz)/dz$? Repetir los cálculos para $\Delta x = 0,001$, $\Delta y = 0,002$. ¿Es la aproximación (en términos relativos) mejor en este segundo caso?
13. Se considera la función $z = f(x, y) = \frac{y}{x}$. Se pide:
- Establecer su dominio de definición.
 - Calcular su gradiente. ¿En qué puntos no existe el gradiente?
 - Obtener la dirección de máximo crecimiento de $f(x, y)$ a partir del punto $(1, 2)$. Si se eligen $\Delta x, \Delta y$ en dicha dirección de forma que se tenga $(\Delta x)^2 + (\Delta y)^2 = 0,1$ ¿cuál es —aproximadamente— el incremento Δz ?
 - A partir de los valores $x = 1, y = 2, z = 2$ ¿cuánto debemos modificar y sin modificar x para que el nuevo valor de z sea aproximadamente $2,01$?
14. Sean $f, g: \mathbb{R}^n \rightarrow \mathbb{R}$ dos funciones diferenciables. Probar que:
- $\nabla(f + g) = \nabla f + \nabla g$
 - $\nabla(fg) = f\nabla g + g\nabla f$
 - $\nabla(f^k) = kf^{k-1}\nabla f$
15. Una ecuación en derivadas parciales de primer orden es una relación que ha de satisfacer una "función incógnita" junto con sus derivadas parciales. Toda función que, sustituida —a la vez que sus derivadas— en la ecuación, convierte a ésta en una identidad en las variables independientes, se dice que es una solución de la ecuación o que satisface la ecuación. Comprobar que las funciones siguientes son soluciones de las ecuaciones correspondientes:

función	ecuación
$z = y \ln(x^2 - y^2)$	$\frac{1}{x} \frac{\partial z}{\partial x} + \frac{1}{y} \frac{\partial z}{\partial y} = \frac{z}{y^2}$
$z = \frac{x^2 + y^2}{\sqrt{x + y}}$	$x \frac{\partial z}{\partial x} + y \frac{\partial z}{\partial y} = \frac{3}{2}z$
$z = \arctan \frac{x^3 + y^3}{x - y}$	$x \frac{\partial z}{\partial x} + y \frac{\partial z}{\partial y} = \sin 2z$
$z = y^2 \operatorname{sen} \frac{y}{x}$	$x^2 \frac{\partial z}{\partial x} + xy \frac{\partial z}{\partial y} = yz$

Capítulo 4

Consecuencias de la diferenciabilidad

La diferenciabilidad es la noción clave sobre la que se desarrolla el cálculo diferencial que extiende a las funciones de varias variables resultados tan fundamentales como la *regla de la cadena* o el *teorema del valor medio*. A ellos dedicamos este capítulo, así como a los primeros pasos en la importante cuestión planteada en el capítulo 1 de si una ecuación $F(x,y) = 0$ define o no una *función implícita*. En el primer apartado se establece la regla de la cadena para derivar funciones compuestas, empezando por la situación más sencilla, en la que queda recogida la idea esencial, y prosiguiendo con sucesivas generalizaciones; como aplicación se ven la fórmula que permite el cálculo de las *derivadas direccionales* de una función diferenciable, el teorema del valor medio para funciones de varias variables con algunas de sus consecuencias, y el teorema de Euler para *funciones homogéneas*, de gran interés en ciertas aplicaciones, en particular en Economía. El segundo apartado está dedicado a la mencionada cuestión de las funciones definidas implícitamente, muy importante en las aplicaciones pues, con frecuencia, el modelo matemático del fenómeno bajo estudio viene dado en forma de ecuaciones que satisfacen las variables implicadas y se plantea saber bajo qué condiciones dichas ecuaciones definen ciertas variables como funciones de las restantes. Estudiaremos esto en general en el capítulo 9.

4.1 La regla de la cadena

La regla de la cadena o de derivación de funciones compuestas de una variable establece que si se tiene la composición de funciones

$$t \xrightarrow{g} x \xrightarrow{f} z \quad (4.1)$$

la derivada de la función compuesta $t \mapsto f(g(t))$, que se denota por $f \circ g$, viene dada por la fórmula $(f \circ g)'(t_0) = f'[g(t_0)] \cdot g'(t_0)$. En notación tradicional, con los riesgos que comporta según se ha comentado en la nota 1 del apartado 3.4, la expresión de la regla de la cadena es extraordinariamente simple:

$$\frac{dz}{dt} = \frac{dz}{dx} \cdot \frac{dx}{dt} \quad (4.2)$$

y parece deducirse de manera inmediata de la existencia de las derivadas dz/dt y dz/dx , pasando al límite cuando $\Delta t \rightarrow 0$ (que implica $\Delta x \rightarrow 0$) en la identidad algebraica

$$\frac{\Delta z}{\Delta t} = \frac{\Delta z}{\Delta x} \cdot \frac{\Delta x}{\Delta t} \quad (4.3)$$

El problema es que la variable $x = g(t)$ puede responder con un incremento $\Delta x = 0$ a un incremento $\Delta t = t - t_0 \neq 0$ y entonces la expresión (4.3) no tiene sentido; por eso la demostración ha de ir por otro camino que, se recordará, utiliza la propiedad de diferenciabilidad de las funciones f y g (o sea, la fórmula (3.5)), que, por ser de una variable, está garantizada por la existencia de derivada.

Parece, pues, claro, que para disponer de una regla de la cadena para funciones de varias variables habrá que partir de la hipótesis de diferenciabilidad. Veamos un primer resultado para funciones de dos variables:

Teorema 4.1 Sea $f(x, y)$ *diferenciable* en un conjunto abierto D de \mathbb{R}^2 y sean $u(t)$ y $v(t)$ *derivables* en un intervalo I de \mathbb{R} tales que $(u(t), v(t)) \in D$ para todo $t \in I$. Entonces, la función compuesta

$$t \mapsto F(t) = f(u(t), v(t)) \quad (4.4)$$

tiene derivada en todo $t \in I$ y vale

$$F'(t) = f_x(u(t), v(t)) \cdot u'(t) + f_y(u(t), v(t)) \cdot v'(t) \quad (4.5)$$

Demostración: Sea $t_0 \in I$ cualquiera. Hay que estudiar si existe y cuánto vale el límite cuando $t \rightarrow t_0$ del cociente incremental

$$\frac{F(t) - F(t_0)}{t - t_0} = \frac{f(u(t), v(t)) - f(u(t_0), v(t_0))}{t - t_0} \quad (4.6)$$

Utilizaremos la diferenciabilidad de f (con la expresión (3.23) del resto):

$$f(x, y) - f(x_0, y_0) = f_x(x_0, y_0) \cdot (x - x_0) + f_y(x_0, y_0) \cdot (y - y_0) + (x - x_0) \cdot \delta_1(x - x_0, y - y_0) + (y - y_0) \cdot \delta_2(x - x_0, y - y_0) \quad (4.7)$$

donde

$$\lim_{(x,y) \rightarrow (x_0,y_0)} \delta_1(x - x_0, y - y_0) = \lim_{(x,y) \rightarrow (x_0,y_0)} \delta_2(x - x_0, y - y_0) = 0 \quad (4.8)$$

en el punto $(x_0, y_0) = (u(t_0), v(t_0))$; con $(x, y) = (u(t), v(t))$ se obtiene de (4.6) y (4.7):

$$\begin{aligned} \frac{F(t) - F(t_0)}{t - t_0} &= f_x(u(t_0), v(t_0)) \cdot \frac{u(t) - u(t_0)}{t - t_0} + \\ &+ f_y(u(t_0), v(t_0)) \cdot \frac{v(t) - v(t_0)}{t - t_0} + \\ &+ \frac{u(t) - u(t_0)}{t - t_0} \cdot \delta_1(u(t) - u(t_0), v(t) - v(t_0)) + \\ &+ \frac{v(t) - v(t_0)}{t - t_0} \cdot \delta_2(u(t) - u(t_0), v(t) - v(t_0)) \end{aligned} \quad (4.9)$$

Como $u(t)$ y $v(t)$ son derivables se tendrá:

$$\frac{u(t) - u(t_0)}{t - t_0} \xrightarrow{t \rightarrow t_0} u'(t_0), \quad \frac{v(t) - v(t_0)}{t - t_0} \xrightarrow{t \rightarrow t_0} v'(t_0)$$

Por otra parte, por ser derivables son continuas (son funciones de una variable), con lo que $(u(t), v(t)) \rightarrow (u(t_0), v(t_0))$ cuando $t \rightarrow t_0$ y, en consecuencia,

$$\lim_{t \rightarrow t_0} \delta_i(u(t) - u(t_0), v(t) - v(t_0)) = 0, \quad i = 1, 2$$

Así pues, tomando límites en (4.9) se tiene:

$$\begin{aligned} \lim_{t \rightarrow t_0} \frac{F(t) - F(t_0)}{t - t_0} &= f_x(u(t_0), v(t_0)) \cdot u'(t_0) + f_y(u(t_0), v(t_0)) \cdot v'(t_0) + \\ &+ u'(t_0) \cdot 0 + v'(t_0) \cdot 0 = f_x(u(t_0), v(t_0)) \cdot u'(t_0) + f_y(u(t_0), v(t_0)) \cdot v'(t_0) \end{aligned}$$

como queríamos demostrar.

Si $z = f(x, y)$, $x = u(t)$, $y = v(t)$, podemos expresar la fórmula (4.5) en la forma abreviada

$$\frac{dz}{dt} = \frac{\partial z}{\partial x} \frac{dx}{dt} + \frac{\partial z}{\partial y} \frac{dy}{dt} \quad (4.10)$$

que, a pesar del riesgo de confusión que entraña el denotar por variables a las funciones y omitir los puntos en que se evalúan éstas, tiene la ventaja de que es mucho más fácil de recordar. Atención: dx y ∂x "no cancelan", su significado es muy diferente y la fórmula (4.10) no tiene la "evidencia algebraica" que tiene (4.2).

EJEMPLO 1. Siendo $z = x^2y^3$, donde $x = e^{2t}$, $y = t^2$, calcúlese dz/dt .

Solución:

$$\frac{dz}{dt} = \frac{\partial z}{\partial x} \frac{dx}{dt} + \frac{\partial z}{\partial y} \frac{dy}{dt} = 2xy^3(2e^{2t}) + 3x^2y^2(2t)$$

Sustituyendo x e y por sus expresiones como funciones de t se obtiene:

$$\frac{dz}{dt} = 4e^{2t}t^6e^{2t} + 6e^{4t}t^4t = 4t^6e^{4t} + 6t^5e^{4t} = 2t^5e^{4t}(2t + 3)$$

En este ejemplo tan simple se puede, naturalmente, construir primero la función compuesta

$$z(t) = (e^{2t})^2(t^2)^3 = t^6e^{4t}$$

y después derivar

$$\frac{dz}{dt} = 6t^5e^{4t} + 4t^6e^{4t} = 2t^5e^{4t}(2t + 3)$$

La demostración del teorema 4.1 se generaliza sin dificultad al caso de funciones de n variables:

Teorema 4.2 Si $f(x_1, \dots, x_n)$ es diferenciable en un conjunto abierto $D \subset \mathbb{R}^n$ y $u_1(t), \dots, u_n(t)$ son funciones derivables en un intervalo I de \mathbb{R} tales que $(u_1(t), \dots, u_n(t)) \in D$ para todo $t \in I$, entonces la función compuesta

$$t \mapsto F(t) = f(u_1(t), \dots, u_n(t)) \quad (4.11)$$

tiene derivada en todo $t \in I$ y vale

$$F'(t) = \frac{\partial f}{\partial x_1}(u_1(t), \dots, u_n(t)) \cdot u_1'(t) + \dots + \frac{\partial f}{\partial x_n}(u_1(t), \dots, u_n(t)) \cdot u_n'(t) \quad (4.12)$$

La versión simplificada de (4.12) es, con $z = f(x_1, \dots, x_n)$ y $x_1 = u_1(t), \dots, x_n = u_n(t)$,

$$\frac{dz}{dt} = \frac{\partial z}{\partial x_1} \frac{dx_1}{dt} + \dots + \frac{\partial z}{\partial x_n} \frac{dx_n}{dt} \quad (4.13)$$

Del teorema 1 se obtiene también una regla de la cadena para el caso en que las variables intermedias dependan a su vez de más de una variable:

Teorema 4.3 Supongamos que $f(x, y)$ es diferenciable en un conjunto abierto $D \subset \mathbb{R}^2$ y que las funciones $u(s, t)$ y $v(s, t)$ tienen derivadas parciales en un conjunto abierto $D' \subset \mathbb{R}^2$. Si se supone que $(u(s, t), v(s, t)) \in D$ para todo $(s, t) \in D'$, entonces la función compuesta

$$(s, t) \mapsto F(s, t) = f(u(s, t), v(s, t)) \quad (4.14)$$

tiene derivadas parciales en D' y valen

$$F_s(s, t) = f_x(u(s, t), v(s, t)) \cdot u_s(s, t) + f_y(u(s, t), v(s, t)) \cdot v_s(s, t) \quad (4.15)$$

$$F_t(s, t) = f_x(u(s, t), v(s, t)) \cdot u_t(s, t) + f_y(u(s, t), v(s, t)) \cdot v_t(s, t) \quad (4.16)$$

Demostración: Para s fijo, $u(s, t)$ y $v(s, t)$ son funciones derivables de la única variable t , y, por el teorema 1, la función $t \mapsto F(s, t)$ es derivable y su derivada, o sea, la derivada parcial $F_t(s, t)$, viene dada por (4.16); la fórmula (4.15) se demuestra de la misma manera.

Como antes, si $z = f(x, y)$ y $x = u(s, t)$, $y = v(s, t)$, podemos expresar (4.15) y (4.16) en la forma simplificada

$$\begin{aligned} \frac{\partial z}{\partial s} &= \frac{\partial z}{\partial x} \frac{\partial x}{\partial s} + \frac{\partial z}{\partial y} \frac{\partial y}{\partial s} \\ \frac{\partial z}{\partial t} &= \frac{\partial z}{\partial x} \frac{\partial x}{\partial t} + \frac{\partial z}{\partial y} \frac{\partial y}{\partial t} \end{aligned}$$

EJEMPLO 2. Si $z = x^2 y^3$ y $x = s^2 + t^2$, $y = st$, se tiene

$$\begin{aligned} \frac{\partial z}{\partial s} &= \frac{\partial z}{\partial x} \frac{\partial x}{\partial s} + \frac{\partial z}{\partial y} \frac{\partial y}{\partial s} = 2xy^3(2s) + 3x^2y^2(t) = \\ &= 4s(s^2 + t^2)s^3t^3 + 3t(s^2 + t^2)^2s^2t^2 = t^3s^2(7s^2 + 3t^2)(s^2 + t^2) \end{aligned}$$

$$\begin{aligned} \frac{\partial z}{\partial t} &= \frac{\partial z}{\partial x} \frac{\partial x}{\partial t} + \frac{\partial z}{\partial y} \frac{\partial y}{\partial t} = 2xy^3(2t) + 3x^2y^2(s) = \\ &= 4t(s^2 + t^2)s^3t^3 + 3s(s^2 + t^2)^2s^2t^2 = s^3t^2(3s^2 + 7t^2)(s^2 + t^2) \end{aligned}$$

EJEMPLO 3. Con frecuencia, es conveniente en el estudio de determinada función $f(x, y)$ hacer un *cambio de variables*, es decir, suponer que x e y dependen de dos nuevas variables

$$x = u(s, t), \quad y = v(s, t)$$

con la idea de que la expresión de la nueva función (compuesta)

$$(s, t) \mapsto F(s, t) = f(u(s, t), v(s, t))$$

sea, para nuestros propósitos, más fácil de manejar que la de f . Si en el problema que nos ocupa intervienen las derivadas de f , interesará también, en ese cambio de variables, expresar las derivadas parciales de una de las funciones, f o F , en términos de las de la otra y de las derivadas de las funciones u y v . Estas relaciones se obtienen mediante la regla de la cadena que da el teorema anterior (y el que sigue, para el caso de más de dos variables). Con la advertencia de rigor, podremos en la práctica denotar con la misma letra z los resultados de las actuaciones de f y F

$$z = f(x, y) = F(s, t)$$

y escribir para las derivadas

$$z_x = f_x, \quad z_y = f_y, \quad z_s = F_s, \quad z_t = F_t$$

Por ejemplo, si en la función $z = f(x, y)$ se hace $x = r \cos \theta$, $y = r \sin \theta$ (cambio a *coordenadas polares*) entonces

$$\begin{aligned} \frac{\partial F}{\partial r}(r, \theta) &= \frac{\partial f}{\partial x}(r \cos \theta, r \sin \theta) \cdot \cos \theta + \frac{\partial f}{\partial y}(r \cos \theta, r \sin \theta) \cdot \sin \theta \\ \frac{\partial F}{\partial \theta}(r, \theta) &= -\frac{\partial f}{\partial x}(r \cos \theta, r \sin \theta) \cdot r \sin \theta + \frac{\partial f}{\partial y}(r \cos \theta, r \sin \theta) \cdot r \cos \theta \end{aligned}$$

y, en forma simplificada,

$$z_r = z_x \cos \theta + z_y \sin \theta, \quad z_\theta = -z_x r \sin \theta + z_y r \cos \theta$$

(Ejercicio: expresar z_x, z_y en términos de z_r, z_θ ; comprobar que se satisface la relación

$$z_x^2 + z_y^2 = z_r^2 + \frac{1}{r^2} z_\theta^2.$$

Esto es la expresión en polares de $|\nabla f|^2$.)

EJEMPLO 4. *Sobre los peligros de las notaciones simplificadas*

Supongamos que se tiene una función $z = f(x, y)$ y que, en la aplicación en que se está trabajando, y resulta ser función de x y de otra variable t : $y = g(x, t)$. Se tiene entonces la función compuesta dependiente de x y t :

$$(x, t) \rightarrow f(x, g(x, t)) = F(x, t)$$

Denotando, como es usual en la práctica, con la misma letra z los resultados de las actuaciones de f y F : $z = f(x, y) = F(x, t)$, hay dos significados para $\frac{\partial z}{\partial x}$. Para no caer en confusión, es necesario indicar cuál es la otra variable independiente que se supone fija en la derivación respecto a x :

$$\left(\frac{\partial z}{\partial x}\right)_y$$

es la derivada respecto a x cuando x e y se consideran independientes e y se mantiene fija (o sea, f_x) mientras que

$$\left(\frac{\partial z}{\partial x}\right)_t$$

es la derivada respecto a x cuando son x y t las variables independientes y t se mantiene fija (o sea, F_x). Con este tipo de notación, la regla de la cadena del teorema anterior da:

$$\left(\frac{\partial z}{\partial x}\right)_t = \left(\frac{\partial z}{\partial x}\right)_y + \left(\frac{\partial z}{\partial y}\right)_x \left(\frac{\partial y}{\partial x}\right)_t \quad (4.17)$$

En efecto, introduciendo provisionalmente una variable s que coincida con x , de modo que se tenga $x = s$, $y = g(s, t)$, o sea, en los términos del enunciado del teorema 4.3, $u(s, t) = s$, $v(s, t) = g(s, t)$, la fórmula (4.17) deriva de (4.15) teniendo en cuenta que $u_s(s, t) = 1$.

Si no se utilizasen los subíndices, la fórmula (4.17) se escribiría

$$\frac{\partial z}{\partial x} = \frac{\partial z}{\partial x} + \frac{\partial z}{\partial y} \frac{\partial y}{\partial x} \quad (4.18)$$

y su interpretación sería confusa, pues, así escrita, implicaría que $\frac{\partial z}{\partial y} \frac{\partial y}{\partial x} = 0$, lo que, naturalmente, no tiene por qué ocurrir en general: por ejemplo, si $z = x + y$ e $y = x + t^2$, se tiene $\frac{\partial z}{\partial y} = 1$, $\frac{\partial y}{\partial x} = 1$ y, en consecuencia, $\frac{\partial z}{\partial y} \frac{\partial y}{\partial x} = 1 \neq 0$.

En Termodinámica, por ejemplo, se hace necesario utilizar este convenio de los subíndices para establecer las relaciones existentes entre las magnitudes básicas que en ella intervienen. Así, la *entalpía* E se puede considerar función

de la presión P y la temperatura T : $E = f(P, T)$, pero, a su vez, la presión es función de la temperatura y el volumen V : $P = g(T, V)$, con lo que

$$\left(\frac{\partial E}{\partial T}\right)_V = \left(\frac{\partial E}{\partial T}\right)_P + \left(\frac{\partial E}{\partial P}\right)_T \left(\frac{\partial P}{\partial T}\right)_V$$

En el cálculo concreto referido a funciones específicas, quizá se puede trabajar más informalmente manteniendo la alerta sobre los peligros señalados; así, si $z = 5x^2 + \sin y$, $y = x^2 + t^2$, haríamos (para la derivada de z respecto a x "final", o sea, de la función compuesta):

$$\begin{aligned} \frac{\partial z}{\partial x} &= \underbrace{15x^2}_{f_z} \cdot \underbrace{1}_{\frac{\partial x}{\partial x}} + \underbrace{\frac{\partial z}{\partial y}}_{f_y} \cdot \frac{\partial y}{\partial x} = 15x^2 + (\cos y)(2x) = \\ &= 15x^2 + 2x \cos y = 15x^2 + 2x \cos(x^2 + t^2) \end{aligned}$$

El teorema 3 se extiende sin dificultad a un número cualquiera de variables:

Teorema 4.4 Supongamos que $f(x_1, \dots, x_n)$ es diferenciable en un conjunto abierto $D \subset \mathbb{R}^n$ y que las funciones $u_1(t_1, \dots, t_m), \dots, u_n(t_1, \dots, t_m)$ tienen derivadas parciales en un conjunto abierto $D' \subset \mathbb{R}^m$. Si se supone que $u_1(t_1, \dots, t_m), \dots, u_n(t_1, \dots, t_m) \in D$ para todo $t = (t_1, \dots, t_m) \in D'$, entonces la función compuesta

$$D' \ni (t_1, \dots, t_m) = t \longmapsto F(t) = f(u_1(t), \dots, u_n(t))$$

tiene derivadas parciales en D' y valen

$$\begin{aligned} \frac{\partial F}{\partial t_1}(t) &= D_1 f(u_1(t), \dots, u_n(t)) \cdot \frac{\partial u_1}{\partial t_1}(t) + \dots + D_n f(u_1(t), \dots, u_n(t)) \cdot \frac{\partial u_n}{\partial t_1}(t) \\ &\quad \dots \dots \dots \\ \frac{\partial F}{\partial t_m}(t) &= D_1 f(u_1(t), \dots, u_n(t)) \cdot \frac{\partial u_1}{\partial t_m}(t) + \dots + D_n f(u_1(t), \dots, u_n(t)) \cdot \frac{\partial u_n}{\partial t_m}(t) \end{aligned} \quad (4.19)$$

En forma simplificada, como en los casos anteriores,

$$\begin{aligned} \frac{\partial z}{\partial t_1} &= \frac{\partial z}{\partial x_1} \frac{\partial x_1}{\partial t_1} + \dots + \frac{\partial z}{\partial x_n} \frac{\partial x_n}{\partial t_1} \\ &\quad \dots \dots \dots \\ \frac{\partial z}{\partial t_m} &= \frac{\partial z}{\partial x_1} \frac{\partial x_1}{\partial t_m} + \dots + \frac{\partial z}{\partial x_n} \frac{\partial x_n}{\partial t_m} \end{aligned} \quad (4.20)$$

o sea

$$\frac{\partial z}{\partial t_j} = \sum_{i=1}^n \frac{\partial z}{\partial x_i} \frac{\partial x_i}{\partial t_j}, \quad j = 1, \dots, m.$$

NOTAS. 1. Si las funciones que se componen son de clase C^1 , es decir, tienen derivadas continuas, las funciones compuestas también son de clase C^1 ; en efecto, las fórmulas que hemos obtenido para las derivadas muestran que éstas son, bajo aquella hipótesis, el resultado de sumas, productos y composiciones de funciones continuas y, por tanto, continuas por los resultados del capítulo 2.

2. En el capítulo 8 veremos el resultado más general de que la composición de funciones (vectoriales, en general) diferenciables es diferenciable y que la diferencial de la función compuesta es la composición de las diferenciales (que son funciones lineales). De este resultado derivarán también estas reglas de la cadena para derivadas parciales que aquí hemos demostrado de manera muy simple.

Derivadas direccionales

De la misma forma que las derivadas parciales de una función $f : D \subseteq \mathbb{R}^2 \rightarrow \mathbb{R}$ representan las pendientes de las curvas coordenadas trazadas sobre la superficie $z = f(x, y)$ siguiendo las direcciones de los ejes x e y , podemos plantearnos, como hicimos en el primer capítulo para el caso de los planos $z = ax + by + c$, qué concepto análogo nos dará la pendiente de la curva trazada sobre la superficie en una *dirección arbitraria* dada por un vector $\vec{u} = (u_1, u_2)$ unitario ($|\vec{u}| = 1$). Como la ecuación paramétrica de la recta que pasa por (x_0, y_0) y tiene a \vec{u} por vector dirección es $\{x = x_0 + tu_1, y = y_0 + tu_2\}$, la mencionada pendiente vendría dada analíticamente por el límite del siguiente cociente incremental:

$$\frac{f(x_0 + tu_1, y_0 + tu_2) - f(x_0, y_0)}{t}$$

obtenido dividiendo por t el incremento que experimenta f al pasar de (x_0, y_0) al punto del plano al que se llega avanzando t unidades con rumbo \vec{u} . Obsérvese que seguimos dentro del esquema general $\Delta f / \Delta s$, donde $\Delta s = t$ representa la distancia (con signo) que separa a los valores finales de las variables independientes de los iniciales, pues

$$\text{dist}((x_0 + tu_1, y_0 + tu_2), (x_0, y_0)) = |(tu_1, tu_2)| = |t| |(u_1, u_2)| = |t|$$

Esto motiva la definición siguiente

Definición 4.1 Sea $f : D \rightarrow \mathbb{R}$, sea (x_0, y_0) un punto interior de D y $\vec{u} = (u_1, u_2)$ un vector unitario. Se llama *derivada de f en (x_0, y_0) en*

la dirección del vector \vec{u} al límite (si existe)

$$\lim_{t \rightarrow 0} \frac{f(x_0 + tu_1, y_0 + tv_2) - f(x_0, y_0)}{t} \quad (4.21)$$

y se denota $D_{\vec{u}}f(x_0, y_0)$ o $\frac{\partial f}{\partial \vec{u}}(x_0, y_0)$.

Siendo α un ángulo entre 0 y 2π , $\alpha \in [0, 2\pi)$, la dirección de ángulo α (o, simplemente, la dirección α) es la dada por el vector unitario $(\cos \alpha, \sin \alpha)$. Se denomina **derivada de f en (x_0, y_0) en la dirección α** a la derivada de f en la dirección del vector $(\cos \alpha, \sin \alpha)$. Se utiliza a veces la representación $D_{\alpha}f(x_0, y_0)$.

Vemos que, a consecuencia de su propia definición, las derivadas parciales no son otra cosa que las derivadas de f en las direcciones principales $\alpha = 0$ ($\vec{u} = \vec{e}_1 = (1, 0)$) y $\alpha = \pi/2$ ($\vec{u} = \vec{e}_2 = (0, 1)$), es decir.

$$D_x f(x_0, y_0) = D_{\vec{e}_1} f(x_0, y_0), \quad D_y f(x_0, y_0) = D_{\vec{e}_2} f(x_0, y_0)$$

Por tanto, las derivadas direccionales son una generalización o extensión del concepto de derivada parcial, y se emplean cuando se desea estudiar el comportamiento incremental de una función cuando sus variables se mueven en una determinada dirección, o lo que es lo mismo, cuando su variación se ve forzada a satisfacer una determinada restricción lineal que no consiste en fijar una de las variables y variar la otra.

EJERCICIO. Probar que $D_{(-\vec{u})}f(x_0, y_0) = -D_{\vec{u}}f(x_0, y_0)$, e interpretar esta fórmula.

Como en el caso de las derivadas parciales, para calcular una derivada direccional basta derivar la función de una variable

$$\phi(t) \stackrel{\text{def}}{=} f(x_0 + tu_1, y_0 + tv_2)$$

particularizando en $t = 0$. Veremos enseguida un método más sencillo para efectuar este cálculo, pero hagamos mientras tanto un ejemplo:

EJEMPLO 5. Para hallar la derivada de $f(x, y) = x^2 - 2xy^3$ en el punto $(1, 2)$ en la dirección de ángulo 30° , obtenemos $\vec{u} = (\cos 30^\circ, \sin 30^\circ) = (\sqrt{3}/2, 1/2)$, y formamos

$$\phi(t) = f(1 + t\sqrt{3}/2, 2 + t/2) = \left(1 + \frac{t\sqrt{3}}{2}\right)^2 - 2\left(1 + \frac{t\sqrt{3}}{2}\right)\left(2 + \frac{t}{2}\right)^3$$

resultando

$$\phi'(t) = \left(1 + \frac{t\sqrt{3}}{2}\right) \sqrt{3} - \sqrt{3} \left(2 + \frac{t}{2}\right)^3 - 3 \left(1 + \frac{t\sqrt{3}}{2}\right) \left(2 + \frac{t}{2}\right)^2$$

y $\phi'(0) = -7\sqrt{3} - 12 \simeq -24.124$. Éste es el valor de la derivada direccional pedida, y significa que, si (x, y) se mueven t unidades de longitud por la recta que pasa por $(1, 2)$ y hace ángulo de 30° con el eje x , la función decrece aproximadamente $-24.124t$ unidades, siendo tanto mejor la aproximación (en términos relativos) cuanto más pequeño sea t . Compárense con una calculadora los valores exacto y aproximado para distintos valores de t .

La interpretación geométrica de las derivadas direccionales es, efectivamente, la que había motivado su definición. En efecto, la gráfica de la función $\phi(t) = f(x_0 + tu_1, y_0 + tu_2)$ no es más que la curva intersección de la superficie $z = f(x, y)$ con el plano vertical cuya ecuación paramétrica es $\{x = x_0 + tu_1, y = y_0 + tu_2, z \text{ arbitrario}\}$, y, por tanto, la derivada de f en (x_0, y_0) según la dirección del vector unitario \vec{u} es la pendiente de la curva intersección de la superficie $z = f(x, y)$ con el plano vertical determinado por la recta del plano $z = 0$ que pasa por (x_0, y_0) y tiene la dirección de \vec{u} (figura 4.1). Llamaremos a ésta "curva coordenada en la dirección \vec{u} " o "curva \vec{u} -coordenada" que pasa por (x_0, y_0, z_0) . Lo mismo que habíamos llamado "paralelo" a la curva coordenada asociada a $\vec{u} = (1, 0)$, y "meridiano" a la asociada a $\vec{u} = (0, 1)$, podríamos llamar "curva de dirección α " a la correspondiente a $\vec{u} = (\cos \alpha, \sin \alpha)$. El vector tangente a esta curva tiene por componentes

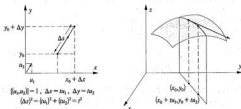


Figura 4.1

horizontales las de \vec{u} , y como componente vertical la propia derivada direccional, es decir: $(u_1, u_2, D_{\vec{u}}f(x_0, y_0))$. De aquí podemos obtener directamente las ecuaciones paramétricas de la recta tangente en la dirección $\vec{u} = (u_1, u_2)$:

$$x = x_0 + tu_1, \quad y = y_0 + tu_2, \quad z = z_0 + tD_{\vec{u}}f(x_0, y_0).$$

Así, por (x_0, y_0, z_0) pasa una familia o "haz" de curvas \vec{u} -coordenadas. Para cada una de estas direcciones, la existencia de la derivada direccional $D_{\vec{u}}f(x_0, y_0)$ equivale a la existencia de recta tangente en (x_0, y_0, z_0) a la curva \vec{u} -coordenada. De éstas, conocemos dos con seguridad: las tangentes al "paralelo" (curva x -coordenada) y al "meridiano" (curva y -coordenada). Si suponemos que f admite derivadas en cualquier dirección, podemos plantearnos qué subconjunto del espacio \mathbb{R}^3 describe este haz de rectas que pasan por (x_0, y_0, z_0) . Lo más natural parece conjeturar que, en general, todas aquellas tangentes se encontrarán en un mismo plano, que no debería ser otro que el determinado por las tangentes al paralelo y al meridiano antes singularizadas ya que no sería lógico que estas direcciones particulares jugaran un papel distinto a las demás. Pues bien, si la función es diferenciable, la respuesta a dicha conjetura es afirmativa: *bajo la hipótesis de diferenciabilidad, las tangentes a las curvas coordenadas en todas las direcciones se hallan en el plano tangente en (x_0, y_0, z_0) a la superficie $z = f(x, y)$* . Esto se deduce de la siguiente fórmula de la derivada direccional, la cual permite, bajo dicha hipótesis, calcular la derivada en cualquier dirección con el mero conocimiento de las derivadas parciales; esto puede parecer sorprendente a primera vista, pero, al fin y al cabo, la esencia de la diferenciabilidad reside, precisamente, en poder calcular —aproximadamente— respuestas a incrementos arbitrarios (h, k) conociendo sólo $f'_x(x_0, y_0)$, $f'_y(x_0, y_0)$, h y k .

Teorema 4.5 (Fórmula de la derivada direccional)

Sean $D \subset \mathbb{R}^2$ abierto, $(x_0, y_0) \in D$, $f : D \rightarrow \mathbb{R}$ diferenciable en (x_0, y_0) y $\vec{u} = (u_1, u_2)$ un vector unitario. Entonces existe $D_{\vec{u}}f(x_0, y_0)$ y vale

$$D_{\vec{u}}f(x_0, y_0) = \frac{\partial f}{\partial x}(x_0, y_0)u_1 + \frac{\partial f}{\partial y}(x_0, y_0)u_2 \quad (4.22)$$

es decir:

$$D_{\vec{u}}f(x_0, y_0) = df(x_0, y_0; u_1, u_2)$$

Por tanto, la derivada de f según el vector \vec{u} coincide con el valor de la diferencial correspondiente a tomar \vec{u} como vector incremento.

Demostración: $D_{\vec{u}}f(x_0, y_0)$ es, por definición, la derivada de la función

$$\phi(t) = f(x_0 + tu_1, y_0 + tu_2)$$

en $t = 0$. Ahora bien, $\phi(t)$ es la función compuesta

$$t \mapsto f(u_1(t), u_2(t))$$

donde $u_1(t) = x_0 + tu_1$, $u_2(t) = y_0 + tu_2$ son funciones de una variable obviamente derivables. Aplicando el teorema 4.1 vemos que $\phi'(t)$ existe y vale:

$$\phi'(t) = f_x(u_1(t), u_2(t)) \cdot u_1 + f_y(u_1(t), u_2(t)) \cdot u_2$$

de donde, para $t = 0$, se obtiene (4.22).

EJEMPLO 6. Calculamos antes, aplicando directamente la definición, la derivada de $f(x, y) = x^2 - 2xy^3$ en el punto $(1, 2)$ en la dirección de ángulo 30° . Recordemos que $\vec{u} = (\cos 30^\circ, \sin 30^\circ) = (\sqrt{3}/2, 1/2)$; por otra parte,

$$\frac{\partial f}{\partial x} = 2x - 2y^3, \quad \frac{\partial f}{\partial y} = -6xy^2$$

de donde

$$\frac{\partial f}{\partial x}(1, 2) = 2 \cdot 1 - 2 \cdot 2^3 = -14, \quad \frac{\partial f}{\partial y}(1, 2) = -6 \cdot 1 \cdot 2^2 = -24$$

y, por tanto,

$$\begin{aligned} D_{\vec{u}}f(1, 2) &= \frac{\partial f}{\partial x}(1, 2)u_1 + \frac{\partial f}{\partial y}(1, 2)u_2 = \\ &= (-14)\frac{\sqrt{3}}{2} + (-24)\frac{1}{2} = -7\sqrt{3} - 12 \end{aligned}$$

que coincide con el valor obtenido anteriormente.

Vamos ahora a comprobar que el plano tangente contiene efectivamente a todas las rectas tangentes a las curvas coordenadas en todas las direcciones. Para ello, nos apoyaremos en la siguiente propiedad, que se deduce directamente de la discusión sobre planos del capítulo 1:

Sea π un plano que pasa por el punto P y tiene vectores directores \vec{V}_1 y \vec{V}_2 . Entonces dicho plano está formado exactamente por todas las rectas que pasan por P y cuyo vector director es combinación lineal de \vec{V}_1 y \vec{V}_2 .

Pues bien, un vector director de la recta tangente en la dirección \vec{u} es, como sabemos,

$$\vec{V}_{\vec{u}} = (u_1, u_2, D_{\vec{u}}f(x_0, y_0))$$

con los casos particulares

$$\left(1, 0, \frac{\partial f}{\partial x}(x_0, y_0)\right) \quad \text{para} \quad \vec{u} = \vec{e}_1, \quad \left(0, 1, \frac{\partial f}{\partial y}(x_0, y_0)\right) \quad \text{para} \quad \vec{u} = \vec{e}_2,$$

para las tangentes en las direcciones coordenadas.

Pero

$$D_{\vec{u}}f(x_0, y_0) = \frac{\partial f}{\partial x}(x_0, y_0)u_1 + \frac{\partial f}{\partial y}(x_0, y_0)u_2$$

y entonces

$$(u_1, u_2, D_{\vec{u}}f(x_0, y_0)) = u_1 \left(1, 0, \frac{\partial f}{\partial x}(x_0, y_0) \right) + u_2 \left(0, 1, \frac{\partial f}{\partial y}(x_0, y_0) \right)$$

y vemos que $\vec{V}_{\vec{u}}$ es combinación lineal de los vectores directores asociados a las direcciones \vec{e}_1 y \vec{e}_2 . Dicho de otra forma, y como queríamos demostrar, la recta tangente en la dirección \vec{u} pasa por (x_0, y_0, z_0) y su vector director está en el plano que contiene a dicho punto y tiene vectores directores $\vec{V}_{\vec{e}_1}$ y $\vec{V}_{\vec{e}_2}$, que no es otro que el plano tangente definido en el capítulo anterior.

De la fórmula de la derivada direccional

$$\begin{aligned} D_{\vec{u}}f(x_0, y_0) &= \frac{\partial f}{\partial x}(x_0, y_0)u_1 + \frac{\partial f}{\partial y}(x_0, y_0)u_2 = \\ &= |\nabla f(x_0, y_0)| \cdot \vec{u} = |\nabla f(x_0, y_0)| \cos \theta \end{aligned}$$

vemos también que la **máxima derivada direccional** se obtiene en la dirección del vector gradiente (cuando $\cos \theta = 1$), o sea, para el vector unitario

$$\frac{\nabla f(x_0, y_0)}{|\nabla f(x_0, y_0)|}$$

y ese valor máximo es $|\nabla f(x_0, y_0)|$. Esto es una reformulación del resultado visto en el capítulo anterior: el gradiente de f en (x_0, y_0) , si no es nulo, marca la dirección de máximo crecimiento de f a partir de (x_0, y_0) , en el sentido de que la pendiente (derivada direccional) en esa dirección es máxima, y el **módulo o norma** del gradiente representa la **máxima derivada direccional**, o sea, el mayor incremento que puede experimentar la función si las variables sufren un incremento (vectorial) unitario —todo esto entendido en *primera aproximación*, es decir, usando la diferencial como aproximación del incremento—. Vemos así que tanto el módulo como la dirección del gradiente tienen una interpretación natural, razón por la cual está justificado considerar el par de derivadas parciales como un vector en su sentido originario de *segmento orientado*.

- NOTAS. 1. Es frecuente también definir la derivada según un vector \vec{u} mediante la fórmula (4.21), sin imponer que \vec{u} sea unitario. Si admitimos ese convenio, es fácil comprobar que, para todo $\alpha \in \mathbb{R}$, $D_{\alpha \vec{u}}f(x_0, y_0) = \alpha D_{\vec{u}}f(x_0, y_0)$ siempre que ésta esté definida. Si bien no hay nada que objetar a esta definición modificada, su sentido geométrico

es distinto, razón por la cual en este libro mantendremos el convenio que estamos siguiendo, es decir, reservar el nombre de derivada direccional a la obtenida mediante (4.21) con un vector unitario. En las notas al final del capítulo haremos más comentarios al respecto.

2. Vimos en el apartado 3.2 mediante el contraejemplo:

$$f(x, y) = \frac{xy^2}{x^2 + y^4}, \quad (x, y) \neq (0, 0) \\ f(0, 0) = 0$$

que, a diferencia de lo que ocurre para funciones de una variable, la existencia de derivadas parciales no implica la continuidad de la función. Compruébese que, de hecho, esta función tiene en $(0, 0)$ derivada según cualquier dirección, por lo que vemos que este grado de "regularidad" a lo largo de cualquier recta que pasa por el origen no garantiza la continuidad en tanto que función de dos variables. Por otro lado, como esta función no es continua, no es diferenciable en $(0, 0)$ y se puede comprobar que las rectas tangentes en ese punto no están contenidas en un mismo plano.

3. A la vista de la fórmula (4.22) se suele preguntar: ¿qué distingue a la derivada direccional de la diferencial, pues ambas tienen la misma forma? La respuesta es que, ciertamente, las derivadas direccionales son particularizaciones de la diferencial, pero no toda evaluación de la diferencial da como resultado una derivada direccional, sino sólo aquellas para las cuales el vector incremento sea unitario. Se da así el hecho curioso de que la diferencial, concebida en un principio sólo para incrementos (h, k) pequeños, se revela como esencial también para incrementos unitarios ($\sqrt{h^2 + k^2} = 1$).

La extensión de lo anterior a funciones de cualquier número de variables no presenta dificultades:

Definición 4.2 Sea $f: D \rightarrow \mathbb{R}$, donde $D \subset \mathbb{R}^n$, y sean $x_0 = (x_1^0, x_2^0, \dots, x_n^0)$ un punto interior de D y $u \in \mathbb{R}^n$ un vector unitario, o sea, de norma $|u| = \sqrt{u_1^2 + u_2^2 + \dots + u_n^2} = 1$. La derivada de f según el vector u en x_0 es el límite, si existe

$$\lim_{t \rightarrow 0} \frac{f(x_1^0 + tu_1, x_2^0 + tu_2, \dots, x_n^0 + tu_n) - f(x_1^0, x_2^0, \dots, x_n^0)}{t} \quad (4.23)$$

es decir,

$$\lim_{t \rightarrow 0} \frac{f(x_0 + tu) - f(x_0)}{t}$$

y se representa por $D_u f(x_0)$.

Teorema 4.6 Sean $f: D \rightarrow \mathbb{R}$, x_0 un punto interior de D , y f diferenciable en x_0 . Si u es un vector unitario, existe la derivada de f en la dirección u y viene dada por la fórmula

$$D_u f(x_0) = \frac{\partial f}{\partial x_1}(x_0)u_1 + \dots + \frac{\partial f}{\partial x_n}(x_0)u_n \quad (4.24)$$

es decir

$$D_u f(x_0) = \nabla f(x_0) \cdot u = df(x_0; u) \quad (4.25)$$

La fórmula de la derivada direccional implica que la recta tangente a la "curva u -coordenada" de ecuaciones paramétricas $\{x = x_0 + tu, z = f(x_0 + tu)\}$ está contenida en el hiperplano tangente.

El teorema del valor medio

El teorema del valor medio (o teorema de los incrementos finitos o teorema de Lagrange) para funciones de una variable: Si $f : [x_0, x_1] \rightarrow \mathbb{R}$ es una función continua y derivable en (x_0, x_1) , existe al menos un $\theta \in (x_0, x_1)$ tal que $f(x_1) - f(x_0) = f'(\theta)(x_1 - x_0)$, admite la siguiente generalización a funciones de varias variables:

Teorema 4.7 (Fórmula de los incrementos finitos)

Sea $f : \mathbb{R}^2 \supset D \rightarrow \mathbb{R}$ y sean $p = (x_0, y_0)$ y $q = (x_1, y_1)$ dos puntos de D tales que el segmento

$$[p, q] = \{(x_0 + t(x_1 - x_0), y_0 + t(y_1 - y_0)); 0 \leq t \leq 1\}$$

que los une está totalmente contenido en D . Si f es diferenciable en todos los puntos de $[p, q]$, entonces

$$\begin{aligned} f(x_1, y_1) - f(x_0, y_0) &= \frac{\partial f}{\partial x}(X, Y) \cdot (x_1 - x_0) + \frac{\partial f}{\partial y}(X, Y) \cdot (y_1 - y_0) = \\ &= df(X, Y; x_1 - x_0, y_1 - y_0) = \nabla f(X, Y) \cdot (x_1 - x_0, y_1 - y_0) \end{aligned} \quad (4.26)$$

para cierto punto intermedio $(X, Y) = (x_0 + \theta(x_1 - x_0), y_0 + \theta(y_1 - y_0))$, $\theta \in (0, 1)$, entre (x_0, y_0) y (x_1, y_1) . Si en la expresión anterior ponemos

$$h = x_1 - x_0, \quad k = y_1 - y_0 \quad (4.27)$$

resulta la formulación equivalente

$$f(x_0 + h, y_0 + k) - f(x_0, y_0) = \frac{\partial f}{\partial x}(X, Y)h + \frac{\partial f}{\partial y}(X, Y)k \quad (4.28)$$

para (x_0, y_0) y (h, k) tales que el segmento que une (x_0, y_0) y $(x_0 + h, y_0 + k)$ está totalmente contenido en D y siendo $(X, Y) = (x_0 + \theta h, y_0 + \theta k)$ para un cierto $\theta \in (0, 1)$.

En efecto, la función de una variable

$$[0, 1] \ni t \mapsto \phi(t) = f(x_0 + th, y_0 + tk) \quad (4.29)$$

verifica

$$\phi(0) = f(x_0, y_0), \quad \phi(1) = f(x_0 + h, y_0 + k)$$

y, por el teorema 4.1, es derivable en $[0, 1]$ y su derivada vale

$$\begin{aligned} \phi'(t) &= f_x(x_0 + th, y_0 + tk) \cdot h + f_y(x_0 + th, y_0 + tk) \cdot k = \\ &= df(x_0 + th, y_0 + tk); h, k = \nabla f(x_0 + th, y_0 + tk) \cdot (h, k) \end{aligned}$$

El teorema resulta, entonces, de aplicar el teorema del valor medio a esta función $\phi(t)$ de una variable: existe $\theta \in (0, 1)$ tal que

$$\phi(1) - \phi(0) = (1 - 0) \cdot \phi'(\theta)$$

utilizando la fórmula que hemos obtenido para $\phi'(\theta)$.

NOTA. Obsérvese que basta suponer que f es continua en todos los puntos de $[p, q]$ y diferenciable en todo punto de $[p, q]$ excepto, posiblemente, los extremos.

En el caso de las funciones de una variable, la consecuencia más inmediata del teorema del valor medio es el importante resultado de que si $f'(x) = 0$ para todo $x \in (a, b)$, entonces f es constante en (a, b) . Para obtener la versión multivariable, recordemos que el análogo multidimensional inmediato de un intervalo es un *conjunto convexo*, es decir, aquél que contiene a todo segmento cuyos extremos pertenezcan a él. Más generalmente, consideremos conjuntos abiertos cuyos puntos puedan unirse por líneas poligonales (una *poligonal* es la unión de un número finito de segmentos S_1, S_2, \dots, S_m que tienen la propiedad de que el extremo final de $S_j = [p_{j-1}, p_j]$ coincide con el inicial de S_{j+1} ; los puntos p_0, p_1, \dots, p_m son los *vértices* de la poligonal y ésta se representa por $[p_0, p_1, \dots, p_m]$):

Definición 4.3 Un conjunto abierto $D \subset \mathbb{R}^2$ se dice que es *convexo* si dados $p, q \in D$, existe una *poligonal* contenida en D que une los puntos p y q , es decir, si existen puntos $p_0, p_1, \dots, p_m \in D$ tales que $p_0 = p$, $p_m = q$, y $[p_{j-1}, p_j] \subset D$, para todo $j = 1, \dots, m$.

Un conjunto abierto convexo es obviamente conexo. La idea intuitiva de conjunto conexo es la de que está "hecho de una sola pieza" (Figura 4.2). En el capítulo 7 profundizaremos en la noción de conexión. Pues bien, se tiene:



Figura 4.2

Proposición 4.1 Sea f una función diferenciable en un conjunto abierto $D \subset \mathbb{R}^2$ conexo. Es condición necesaria y suficiente para que f sea constante en D el que

$$\frac{\partial f}{\partial x}(x, y) = \frac{\partial f}{\partial y}(x, y) = 0 \quad (4.30)$$

para todo $(x, y) \in D$.

La necesidad es obvia, y en cuanto a la suficiencia, fijemos $p = (x_0, y_0) \in D$ y sea $q = (x, y) \in D$ arbitrario. Por hipótesis, existe una poligonal $[p = p_0, p_1, \dots, p_m = q]$, $p_j = (x_j, y_j)$, $j = 0, 1, \dots, m$, que los une y está enteramente contenida en D . Aplicando en cada uno de los intervalos que la componen el teorema del valor medio se tiene $f(x_{j-1}, y_{j-1}) = f(x_j, y_j)$ para todo $j = 1, \dots, m$, y, en consecuencia $f(x_0, y_0) = f(x_m, y_m)$, es decir, $f(x_0, y_0) = f(x, y)$.

Corolario 4.1 Sean f y g dos funciones diferenciables en un conjunto abierto D conexo tales que $\nabla f(x, y) = \nabla g(x, y)$ para todo $(x, y) \in D$. Entonces, existe una constante c tal que $f(x, y) = g(x, y) + c$ para todo $x \in D$, o sea, f y g difieren en una constante.

En efecto, basta aplicar la proposición anterior a la diferencia $f - g$. (Resaltamos que la restricción sobre la geometría de D para la validez de este resultado ya era decisiva en las funciones de una variable: no es cierto, en general, que si $f'(x) = g'(x)$ entonces f y g difieren en una constante; para garantizarlo, hay que suponer que f y g están definidas en cierto intervalo I ; por ejemplo, las funciones

$$f(x) = \begin{cases} \frac{1}{x} - 1 & \text{si } x < 0 \\ \frac{1}{x} + 1 & \text{si } x > 0 \end{cases}$$

$$g(x) = \frac{1}{x} \quad \text{si } x \neq 0$$

tienen la misma derivada en el dominio común $\mathbb{R} \setminus \{0\}$ y no existe una constante c tal que $f(x) = g(x) + c$. En \mathbb{R} , los conjuntos abiertos conexos son los intervalos abiertos, como veremos en el capítulo 7.)

Si $f(x, y)$ es tal que existe una función g de una variable de modo que $f(x, y) = g(x)$ para todo $(x, y) \in D$, entonces se dice que f *no depende* de y , y es claro que $f'_y(x, y) = 0$ para todo $(x, y) \in D$. Dada una función diferenciable f en un conjunto abierto conexo D ¿será suficiente la condición $f'_y(x, y) = 0$ para concluir que f no depende de y ? La respuesta es *no*, como muestra el ejemplo de la función f definida en $D = \mathbb{R}^2 \setminus \{(x, 0) : x \leq 0\}$ por

$$f(x, y) = \begin{cases} x^2 & \text{si } x > 0 \text{ o } y > 0 \\ -x^2 & \text{si } x \leq 0 \text{ e } y < 0 \end{cases}$$

La función es diferenciable en D , verifica $f'_y(x, y) = 0$ y, sin embargo, $f(-1, -1) = -1$ y $f(-1, 1) = 1$. El "problema" está, en realidad, en la geometría del conjunto D , que sí es conexo pero no "convexo según verticales": compruébese que si D es tal que para cada par de puntos $(x, y_1), (x, y_2) \in D$ el segmento (vertical) que los une está completamente contenido en D , entonces la condición $f'_y(x, y) = 0$ sí es suficiente para que f no dependa de y . Naturalmente, se tiene el resultado análogo para la variable x (considerando segmentos horizontales). En un conjunto *convexo* (por ejemplo, todo el plano, un círculo,...) siempre se podrá concluir de la anulación en todos sus puntos de la derivada respecto a una variable que la función (supuesta diferenciable) no depende de esa variable.

Dado un abierto convexo D (el plano entero, por ejemplo) ¿cómo serán las funciones de clase $C^1(D)$ tales que $f_x = 0$? Según lo anterior, serán de la forma $f = \varphi(y)$ con φ una función arbitraria (de clase 1) de una variable. En otras palabras, el conjunto de soluciones (o solución general) de la ecuación en derivadas parciales $u_x = 0$ está dado por $u(x, y) = \varphi(y)$ con φ una función arbitraria. Si nos hacemos la misma pregunta para la ecuación

$$u_x = x^2 y$$

basta observar que, como

$$\frac{\partial}{\partial x} \left(\frac{x^3}{3} y \right) = x^2 y$$

ésta es equivalente a

$$\frac{\partial}{\partial x} \left(u - \frac{x^3}{3} y \right) = 0$$

de donde se deduce que $u - \frac{x^3}{3}y = \varphi(y)$, es decir, que las soluciones son

$$u(x, y) = \frac{x^3}{3}y + \varphi(y)$$

con $\varphi(y)$ arbitraria. La manera de proceder en la práctica consiste en utilizar el cálculo integral de una variable, integrando, en este caso, respecto a x —tratando a y como constante— y teniendo en cuenta que ahora la “constante de integración” es una función de y (o de x si se trata de una integración respecto a y):

$$u(x, y) = \int x^2 y \, dx + \varphi(y) = \frac{x^3 y}{3} + \varphi(y)$$

EJERCICIO Resuélvanse las ecuaciones

$$\begin{array}{ll} a) u_x = h(y) & b) u_y = g(x) \\ c) u_x = g(x) + h(y) & d) u_y = g(x) + h(y) \end{array}$$

(siendo g y h funciones conocidas continuas en todo \mathbb{R}).

El teorema del valor medio para funciones de n variables es:

Teorema 4.8 Sean $a, b \in D$, D subconjunto abierto de \mathbb{R}^n , tales que el segmento $[a, b] = \{a + t(b - a) : 0 \leq t \leq 1\}$ está contenido en D . Si $f : D \rightarrow \mathbb{R}$ es diferenciable en todos los puntos de $[a, b]$, entonces existe $\theta \in (0, 1)$ tal que

$$f(b) - f(a) = \nabla f(a + \theta(b - a)) \cdot (b - a)$$

La demostración es la misma que para funciones de dos variables, así como sus consecuencias relativas a funciones diferenciables en conjuntos abiertos conexos, cuya definición es la misma que en \mathbb{R}^2 utilizando segmentos en el espacio n -dimensional:

Proposición 4.2 Sea f diferenciable en un conjunto abierto $D \subset \mathbb{R}^n$ conexo. Es condición necesaria y suficiente para que f sea constante en D el que

$$\frac{\partial f}{\partial x_1}(x) = \dots = \frac{\partial f}{\partial x_n}(x) = 0 \quad (4.31)$$

para todo $x \in D$.

Corolario 4.2 Sean f y g dos funciones diferenciables en un conjunto abierto D conexo tales que $\nabla f(x) = \nabla g(x)$ para todo $x \in D$. Entonces, existe una constante c tal que $f(x) = g(x) + c$ para todo $x \in D$, o sea, f y g difieren en una constante.

Reseñemos también otra consecuencia inmediata del teorema:

Corolario 4.3 Si f es diferenciable en un subconjunto convexo C y existe una constante k tal que $|\nabla f(x)| \leq k$ para todo $x \in C$, entonces

$$|f(x) - f(y)| \leq k|x - y| \quad (4.32)$$

para todo par de puntos $x, y \in C$ (o sea, f es lipechitziana en C).

Finalmente, si f es diferenciable en un conjunto convexo C y $D_i f(x) = 0$ para todo $x \in C$, entonces f es independiente de x_i , es decir, existe una función g de $n-1$ variables tal que $f(x_1, \dots, x_n) = g(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n)$ para todo $x = (x_1, \dots, x_n) \in C$.

Funciones homogéneas

Una función $f(x)$ definida en un conjunto abierto $D \subset \mathbb{R}^n$ se dice que es *homogénea de grado k en D* si para todo $x \in D$ se verifica

$$f(tx_1, \dots, tx_n) = t^k f(x_1, \dots, x_n) \quad (4.33)$$

para todo número real $t > 0$ suficientemente próximo a 1

(Obsérvese que, al ser D abierto, si $x \in D$ entonces $tx \in D$ para t suficientemente próximo a 1.)

Por ejemplo, las funciones

$$x^3 + 3x^2y + 5y^3, \quad \frac{x^2 - y^2}{x^2 + y^2}, \quad \frac{1}{\sqrt{x^2 + y^2 + z^2}}$$

son homogéneas de grados 3, 0 y -1, respectivamente. La función $f(x, y) = cx^ay^b$, con a, b, c constantes, definida en $\{(x, y); x > 0, y > 0\}$, es homogénea de grado $a+b$. Si $a+b=1$ se trata de una función de producción de Cobb-Douglas.

Usualmente, se supone que las funciones homogéneas están definidas en un cono de vértice el origen, o sea, un conjunto tal que $x \in D$ y $t > 0$ implican que $tx = (tx_1, \dots, tx_n) \in D$ y que se verifica (4.33) para todo $t > 0$ (en los ejemplos anteriores, los "conos" son, respectivamente, \mathbb{R}^2 , $\mathbb{R}^2 \setminus \{(0, 0)\}$, $\mathbb{R}^3 \setminus \{(0, 0, 0)\}$ y el primer cuadrante abierto del plano $\{(x, y); x > 0, y > 0\}$). Obsérvese que, en esta situación, los valores de f quedan determinados por los valores que toma en (la intersección de D con) la esfera unidad, pues, para cualquier $x \in D$, $x \neq 0$, el vector

$$u(x) = \frac{1}{|x|}x$$

tiene norma 1, pertenece a D , y, tomando $t = |x|$, se tiene

$$f(x_1, \dots, x_n) = f\left(t \frac{x_1}{|x|}, \dots, t \frac{x_n}{|x|}\right) = t^k f\left(\frac{x}{|x|}\right)$$

(fig. 4.3; si el origen pertenece a D , compruébese que, si $k \neq 0$, ha de ser $f(0, \dots, 0) = 0$, y que, si $k = 0$ y se supone que f es continua en el origen, entonces f es constante. Indicación: $f(0) = \lim_{t \rightarrow 0^+} f(t\bar{x})$ para todo $\bar{x} \neq 0$).

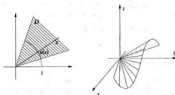


Figura 4.3

En realidad, basta conocer el valor de la función en un punto de cada rayo que parte del origen (o sea, cada semirrecta de la forma $x = t\bar{x}$ con $t > 0$ y $\bar{x} \in D$). Para una función de dos variables $z = f(x, y)$, los valores sobre cada rayo $(t\bar{x}, t\bar{y})$, (\bar{x}, \bar{y}) fijado, están dados por $z = t^k \bar{z}$, donde $\bar{z} = f(\bar{x}, \bar{y})$ y t representa la distancia al origen a lo largo del rayo a través de la relación $|(t\bar{x}, t\bar{y})| = t |(\bar{x}, \bar{y})|$; si, en particular, la función es homogénea de grado 1, entonces $z = t\bar{z}$, es decir, la curva que resulta al cortar la gráfica de $z = f(x, y)$ por el semiplano vertical que contiene al rayo $(t\bar{x}, t\bar{y})$ es una recta y así resulta que la gráfica de una función homogénea de grado 1 es una superficie engendrada por rectas en el espacio que pasan por el origen (Figura 4.3).

EJERCICIO. Compruébese que si $f \in C^1(D)$ es homogénea de grado k , entonces sus derivadas $D_i f$ son funciones homogéneas de grado $k - 1$.

Las funciones homogéneas diferenciables quedan caracterizadas, gracias a la regla de la cadena, por el siguiente resultado:

Teorema 4.9 (Teorema de Euler) Una función diferenciable en D es homogénea de grado k si y sólo si

$$x_1 D_1 f(x_1, \dots, x_n) + \dots + x_n D_n f(x_1, \dots, x_n) = k f(x_1, \dots, x_n) \quad (4.34)$$

para todo $x \in D$.

Demostración: Consideremos $f(tx_1, \dots, tx_n)$ como función compuesta de t :

$$(1 - \varepsilon, 1 + \varepsilon) \ni t \mapsto f(tx_1, \dots, tx_n)$$

para cada $(x_1, \dots, x_n) \in D$ fijo. Derivando la ecuación de (4.33) respecto a t resulta, por la regla de la cadena,

$$x_1 D_1 f(tx_1, \dots, tx_n) + \dots + x_n D_n f(tx_1, \dots, tx_n) = kt^{k-1} f(x_1, \dots, x_n)$$

de donde, para $t = 1$, se obtiene (4.34).

Recíprocamente, supongamos que (4.34) es cierta para todo $x \in D$. Fijado x , definamos la función

$$g(t) = \frac{f(tx_1, \dots, tx_n)}{t^k}$$

De la verificación de (4.34) en (tx_1, \dots, tx_n) para valores de t suficientemente próximos a 1:

$$(tx_1) D_1 f(tx_1, \dots, tx_n) + \dots + (tx_n) D_n f(tx_1, \dots, tx_n) = k f(tx_1, \dots, tx_n)$$

se deduce que

$$\begin{aligned} g'(t) &= \frac{t^k \sum_{i=1}^n x_i D_i f(tx_1, \dots, tx_n) - kt^{k-1} f(tx_1, \dots, tx_n)}{t^{2k}} = \\ &= \frac{t^{k-1} [\sum_{i=1}^n tx_i D_i f(tx_1, \dots, tx_n) - k f(tx_1, \dots, tx_n)]}{t^{2k}} = 0 \end{aligned}$$

es decir, g es constante en t ; se tiene, por lo tanto, $g(t) = g(1)$, pero esto es, por la definición de g , la propiedad (4.33) que define a las funciones homogéneas.

La regla de Leibniz

También se puede utilizar la regla de la cadena en la demostración de la fórmula de Leibniz para derivar integrales que dependen de un parámetro, cuestión importante en diversas aplicaciones. El lector conocerá del curso de funciones de una variable que, de acuerdo con el *teorema fundamental del cálculo*,

$$\frac{d}{dt} \int_a^t f(x) dx = f(t) \quad (4.35)$$

y, teniendo en cuenta que $\int_t^b f(x) dx = -\int_b^t f(x) dx$,

$$\frac{d}{dt} \int_t^b f(x) dx = -f(t) \quad (4.36)$$

Más generalmente, si $a(t)$ y $b(t)$ son derivables y $f(x)$ es continua se tiene

$$\frac{d}{dt} \int_{a(t)}^{b(t)} f(x) dx = f(b(t))b'(t) - f(a(t))a'(t) \quad (4.37)$$

(para la demostración, se parte de

$$\int_{a(t)}^{b(t)} f(x) dx = F(b(t)) - F(a(t))$$

siendo F una primitiva de f , y se deriva respecto a t aplicando la regla de la cadena).

La regla de Leibniz generaliza el resultado (4.37) al caso en que tanto los límites de integración como el integrando dependen del parámetro t : supongamos que $f(t, x)$ es una función continua en el rectángulo

$$R = \{(t, x) \in \mathbb{R}^2; t \in I, a \leq x \leq b\}$$

donde I es un intervalo de la recta real, y que $a(t)$ y $b(t)$ son funciones continuas definidas en I tales que $a \leq a(t) \leq b(t) \leq b$ para todo $t \in I$ (figura 4.4).

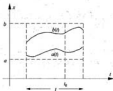


Figura 4.4

Para cada $t_0 \in I$ se tiene una función continua de una variable

$$x \mapsto f(t_0, x)$$

definida en el intervalo $[a(t_0), b(t_0)]$; la correspondiente integral definida

$$\int_{a(t_0)}^{b(t_0)} f(t_0, x) dx$$

es un número que depende de t_0 ; se tiene entonces definida en I la función:

$$G(t) = \int_{a(t)}^{b(t)} f(t, x) dx$$

que se llama *integral dependiente de un parámetro entre los límites variables* $a(t)$ y $b(t)$. Pues bien, se tiene el siguiente

Teorema 4.10 (Regla de Leibniz) Si $a(t)$ y $b(t)$ son derivables y $f'_t(t, x)$ existe y es continua en R , entonces $G(t)$ es derivable y su derivada está dada por la fórmula

$$G'(t) = f(t, b(t))b'(t) - f(t, a(t))a'(t) + \int_{a(t)}^{b(t)} \frac{\partial f(t, x)}{\partial t} dx \quad (4.38)$$

La fórmula (4.37) es obviamente un caso particular de ésta cuando f no depende de t . A su vez, cuando los límites de integración son fijos resulta

$$G(t) = \int_a^b f(t, x) dx$$

y

$$\frac{d}{dt} \int_a^b f(t, x) dx = \int_a^b \frac{\partial f(t, x)}{\partial t} dx \quad (4.39)$$

(la derivada de la integral dependiente de un parámetro se obtiene derivando bajo el signo integral, si bien hay que cambiar d/dt por $\partial/\partial t$ pues esta última derivación afecta a una función de dos variables).

La idea para demostrar (4.38) es la siguiente: se considera la función de tres variables

$$H(t, u, v) = \int_u^v f(t, x) dx$$

donde $t \in I$ y $u, v \in [a, b]$. Una vez demostrado que H es diferenciable, se aplica la regla de la cadena a la función compuesta

$$t \mapsto H(t, a(t), b(t)) = G(t) :$$

$$G'(t) = H'_t(t, a(t), b(t)) + H'_u(t, a(t), b(t))a'(t) + H'_v(t, a(t), b(t))b'(t)$$

Por (4.35) y (4.36) se tiene que

$$H'_u = -f(t, u), \quad H'_v = f(t, v)$$

que, por la hipótesis, son funciones continuas en $t \in I$, $u, v \in [a, b]$. Así que la única dificultad para obtener (4.38) estriba en demostrar que

$$H'_t = \int_u^v \frac{\partial f(t, x)}{\partial t} dx \quad (4.40)$$

es decir, la fórmula (4.39); si además se prueba que $H'_t(t, u, v)$ es continua en (t, u, v) , entonces H es diferenciable (pues sus derivadas parciales son continuas) y es legítima la anterior aplicación de la regla de la cadena para obtener (4.38). Demostraremos esto al final de este párrafo. Veamos antes algunos ejemplos:

1. Sea

$$G(t) = \int_0^1 \cos(t^2 + x) dx$$

Por (4.39)

$$\begin{aligned} G'(t) &= \int_0^1 -2t \operatorname{sen}(t^2 + x) dx = 2t \int_0^1 -\operatorname{sen}(t^2 + x) dx = \\ &= 2t [\cos(t^2 + x)]_0^1 = 2t [\cos(t^2 + 1) - \cos t^2] \end{aligned}$$

En este caso, se puede obtener $G'(t)$ calculando explícitamente $G(t)$ por integración elemental:

$$G(t) = \int_0^1 \cos(t^2 + x) dx = \operatorname{sen}(t^2 + 1) - \operatorname{sen} t^2$$

y derivando después respecto a t : $G'(t) = 2t [\cos(t^2 + 1) - \cos t^2]$

2. En otros casos, la regla de Leibniz es el único medio para obtener explícitamente $G'(t)$: sea, por ejemplo,

$$G(t) = \int_1^2 \frac{e^{tx}}{x} dx$$

Por (4.39)

$$G'(t) = \int_1^2 x \frac{e^{tx}}{x} dx = \int_1^2 e^{tx} dx = \left[\frac{e^{tx}}{t} \right]_1^2 = \frac{e^{2t} - e^t}{t}$$

En este caso no es posible calcular explícitamente $G(t)$ pues la primitiva de e^{tx}/x no es una función elemental.

3. Sea

$$G(t) = \int_t^{t^2} \operatorname{sen}(t^2 + x^2) dx$$

Por (4.38)

$$G'(t) = 2t \cdot \operatorname{sen}(t^2 + t^4) - 1 \cdot \operatorname{sen}(t^2 + x^2) + \int_t^{t^2} 2t \cos(t^2 + x^2) dx$$

NOTA. En cuanto a la demostración pendiente —que se puede pasar por alto en una primera lectura— probemos en primer lugar que, si $f(t, x)$ es continua en el rectángulo $R = \{(t, x) \in \mathbb{R}^2; t \in I, a \leq x \leq b\}$, entonces

$$H(t, u, v) = \int_u^v f(t, x) dx$$

es continua en $B = \{(t, u, v) \in \mathbb{R}^3; t \in I, a \leq u, v \leq b\}$ (resultado interesante por sí mismo sobre la continuidad de la integral dependiente de un parámetro, antes de plantearse la cuestión de la derivabilidad). Dado $(t_0, u_0, v_0) \in B$, sean h, k, l suficientemente pequeños para que también $(t_0 + h, u_0 + k, v_0 + l) \in B$. Se tiene:

$$\begin{aligned} & H(t_0 + h, u_0 + k, v_0 + l) - H(t_0, u_0, v_0) = \\ &= \int_{u_0+k}^{v_0+l} f(t_0 + h, x) dx - \int_{u_0}^{v_0} f(t_0, x) dx = \\ &= \int_{u_0+k}^{v_0+l} f(t_0 + h, x) dx - \int_{u_0+k}^{v_0+l} f(t_0, x) dx + \\ &+ \int_{u_0+k}^{v_0+l} f(t_0, x) dx - \int_{u_0}^{v_0} f(t_0, x) dx = \\ &= \int_{u_0+k}^{v_0+l} [f(t_0 + h, x) - f(t_0, x)] dx + \int_{u_0+k}^{v_0+l} f(t_0, x) dx - \\ &- \int_{u_0+k}^{v_0} f(t_0, x) dx - \int_{u_0}^{v_0} f(t_0, x) dx = \\ &= \int_{u_0+k}^{v_0+l} [f(t_0 + h, x) - f(t_0, x)] dx + \int_{v_0}^{v_0+l} f(t_0, x) dx - \int_{u_0}^{u_0+k} f(t_0, x) dx \end{aligned}$$

Para ρ suficientemente pequeño, el rectángulo

$$R_\rho = \{(t, x) \in \mathbb{R}^2; |t - t_0| \leq \rho, a \leq x \leq b\}$$

es un conjunto cerrado y acotado contenido en R en el que f es, como dijimos en el capítulo 2 y demostraremos en el capítulo 7, uniformemente continua. En consecuencia, para todo $\varepsilon > 0$ existe un δ_1 , $0 < \delta_1 \leq \rho$, tal que

$$|f(t_0 + h, x) - f(t_0, x)| \leq \varepsilon/3(b-a)$$

para todo $x \in [a, b]$ y todo h tal que $|h| \leq \delta_1$. Por otro lado, si $M(t_0)$ es el máximo de la función continua de una variable $x \mapsto |f(t_0, x)|$ en el intervalo $[a, b]$, se tiene:

$$\left| \int_{u_0}^{u_0+k} f(t_0, x) dx \right| \leq M(t_0) |k| \leq \frac{\varepsilon}{3}, \quad \left| \int_{v_0}^{v_0+l} f(t_0, x) dx \right| \leq M(t_0) |l| \leq \frac{\varepsilon}{3}$$

eligiendo k y l con $|k|, |l| \leq \varepsilon/3M(t_0)$. Tomando $\delta = \min(\delta_1, \varepsilon/3M(t_0))$, vemos que, si

$$(h^2 + k^2 + l^2)^{1/2} \leq \delta$$

entonces $|h|, |k|, |l| \leq \delta$ y

$$|H(t_0 + h, u_0 + k, v_0 + l) - H(t_0, u_0, v_0)| \leq \frac{\varepsilon}{3(b-a)}(b-a) + \varepsilon/3 + \varepsilon/3 = \varepsilon$$

quedando demostrada la continuidad de H . Para demostrar (4.40) escribimos (con u y v fijados)

$$\begin{aligned} & \frac{H(t_0 + h, u, v) - H(t_0, u, v)}{h} = \int_u^v f'_t(t, x) dx = \\ &= \frac{1}{h} \int_u^v [f(t_0 + h, x) - f(t_0, x)] dx = \int_u^v f'_t(t, x) dx = \\ &= (\text{por el teorema del valor medio}) = \\ &= \int_u^v [f'_t(t_0 + \theta h, x) - f'_t(t, x)] dx \end{aligned}$$

Por el mismo argumento que antes, utilizando ahora la continuidad de $f'_t(t, x)$, esta expresión converge a cero cuando $h \rightarrow 0$, es decir

$$H'_t(t_0, u, v) = \lim_{h \rightarrow 0} \frac{H(t_0 + h, u, v) - H(t_0, u, v)}{h} = \int_u^v \frac{\partial f(t, x)}{\partial t} dx$$

y, también por el mismo argumento, se tiene que

$$H'_t(t, u, v) = \int_u^v \frac{\partial f(t, x)}{\partial t} dx$$

es continua en (t, u, v) .

4.2 Funciones implícitas

Como decíamos en el apartado 1.2, la ecuación en dos variables

$$F(x, y) = 0 \tag{4.41}$$

define implícitamente a la función $y = f(x)$ en un intervalo $I \subseteq \mathbb{R}$ si

$$F(x, f(x)) = 0 \tag{4.42}$$

para todo $x \in I$. Consideremos algunos ejemplos:

1. La ecuación

$$x^2 + y^2 + 1 = 0 \tag{4.43}$$

no se satisface en ningún punto (x, y) del plano y en consecuencia no define ninguna función $y = f(x)$

2. La ecuación

$$x^2 + y^2 = 0 \quad (4.44)$$

sólo se satisface en el origen $(0, 0) \in \mathbb{R}^2$, con lo que no definirá implícitamente ninguna función significativa.

3. La ecuación

$$x^3 + y^3 - 1 = 0 \quad (4.45)$$

sí define a y como función implícita de x , y en este caso es posible expresar a y como función explícita de x :

$$y = \sqrt[3]{1 - x^3}, \quad x \in \mathbb{R} \quad (4.46)$$

4. En la ecuación

$$x^2 + y^2 - 1 = 0 \quad (4.47)$$

también es posible "despejar" y como función de x :

$$y = \pm \sqrt{1 - x^2}, \quad x \in [-1, 1] \quad (4.48)$$

Ahora, la ecuación (4.47) define *dos* funciones $y = f(x)$ (si se impone la exigencia de que $y = f(x)$ ha de ser continua; si no, se tienen definidas infinitas funciones).

$$y = \sigma(x) \sqrt{1 - x^2}$$

con $\sigma(x) = 1$ ó -1 , elegido arbitrariamente, para cada $x \in [-1, 1]$. Obsérvese, sin embargo, que (4.47) define implícitamente una *única* función (continua) $y = f(x)$ en el entorno de cada punto (x_0, y_0) que satisfaga la ecuación (4.47) y sea tal que $y_0 \neq 0$; y en el entorno de los puntos $(-1, 0)$ y $(1, 0)$, (4.47) define una única función $x = g(y)$.

5. En la ecuación

$$x^3 + xy + y^3 = 0$$

sería posible, aunque complicado algebraicamente, despejar y como función de x (utilizando las fórmulas que dan las raíces de una ecuación de tercer grado).

6. La ecuación

$$y^5 - y^2 + 2xy - x^2 = 0 \quad (4.49)$$

define a y como función implícita de x : para cada $x \in \mathbb{R}$, (4.49) es una ecuación de grado impar en y que tiene al menos una solución. En este caso no es posible expresar y explícitamente en términos de funciones elementales de x (no es posible, en general, expresar las raíces de una ecuación de quinto grado, o superior, mediante expresiones radicales de sus coeficientes, o sea, expresiones compuestas de las cuatro operaciones racionales y extracción de raíces de índices cualesquiera).

7. En la ecuación

$$x - 2 \ln x + y - 2 \ln y - 2 = 0 \quad (4.50)$$

no es posible despejar la variable y y expresarla explícitamente en términos de funciones elementales de la variable x (ni tampoco x en términos de y), pero, como veremos enseguida, (4.50) define implícitamente a y como función de x (y a x como función de y).

Vemos, pues, que una ecuación $F(x, y) = 0$ puede o no definir implícitamente una función $y = f(x)$, y en caso de hacerlo puede o no ser posible expresar y en función de x explícitamente en términos de funciones elementales. Esta cuestión puede contemplarse desde otro punto de vista equivalente que parte del interés en resolver esa ecuación en dos variables: supuesto que se conoce una solución (x_0, y_0) , $F(x_0, y_0) = 0$, interesa saber si existen otras soluciones de la ecuación, o sea, si para $x \neq x_0$, acaso con la restricción de ser próximo a x_0 , existe $y = f(x)$ de modo que $F(x, f(x)) = 0$; se tendría entonces toda una curva $y = f(x)$ de soluciones de la ecuación en dos variables (la hipótesis de que la ecuación tiene al menos una solución a partir de la cual obtener otras parece razonable: considérese el primero de los ejemplos anteriores). Si $F(x, y)$ es diferenciable, se puede escribir, como sabemos,

$$\begin{aligned} 0 &= F(x, y) = F(x, y) - F(x_0, y_0) = \\ &= \frac{\partial F}{\partial x}(x_0, y_0)(x - x_0) + \frac{\partial F}{\partial y}(x_0, y_0)(y - y_0) + G(x, y, x_0, y_0) \end{aligned} \quad (4.51)$$

con

$$\lim_{(x,y) \rightarrow (x_0,y_0)} \frac{G(x_0, y_0, x, y)}{\sqrt{(x - x_0)^2 + (y - y_0)^2}} = 0$$

La idea para tratar de resolver la ecuación $F(x, y) = 0$ es, una vez más, sustituir F por su *aproximación lineal* y conjeturar, dada la "pequeñez" de

G en las proximidades de (x_0, y_0) , que los resultados que se obtengan para la ecuación linealizada

$$\frac{\partial F}{\partial x}(x_0, y_0)(x - x_0) + \frac{\partial F}{\partial y}(x_0, y_0)(y - y_0) = 0 \quad (4.52)$$

sean aplicables a la ecuación (4.51) en un entorno suficientemente pequeño de (x_0, y_0) .

Dado que en (4.52) y está definida (explícitamente) como función de x si

$$\frac{\partial F}{\partial y}(x_0, y_0) \neq 0 \quad (4.53)$$

$$\left(y = y_0 - \left[\frac{\partial F}{\partial y}(x_0, y_0) \right]^{-1} \left[\frac{\partial F}{\partial x}(x_0, y_0)(x - x_0) \right] \right)$$

parece claro que la hipótesis clave para poder hacer lo mismo cerca de (x_0, y_0) en la ecuación no lineal (4.51) (al menos teóricamente) ha de ser (4.53).

Efectivamente, así es, y se tiene el siguiente resultado fundamental que demostraremos en el capítulo 9:

Teorema 4.11 (Teorema de la función implícita)

Si $F(x, y)$ es una función con derivadas continuas en un conjunto abierto $D \subseteq \mathbb{R}^2$ y $(x_0, y_0) \in D$ es tal que $F(x_0, y_0) = 0$ y $F'_y(x_0, y_0) \neq 0$, entonces existen $\delta > 0, \varepsilon > 0$ tales que

i) Para cada $x \in (x_0 - \delta, x_0 + \delta)$, la ecuación $F(x, y) = 0$ tiene una única solución en $(y_0 - \varepsilon, y_0 + \varepsilon)$.

ii) Denotando esa única solución por $f(x)$, la función $y = f(x)$ tiene derivada continua en $(x_0 - \delta, x_0 + \delta)$ dada por

$$f'(x) = -\frac{F'_x(x, f(x))}{F'_y(x, f(x))} \quad (4.54)$$

Obsérvese que $y_0 = f(x_0)$, por la unicidad de solución en $(x_0 - \delta, x_0 + \delta)$; se tiene, pues, una curva de soluciones de $F(x, y) = 0$ que pasa por (x_0, y_0) . Las hipótesis de continuidad de F'_y y $F'_y(x_0, y_0) \neq 0$ implican que $F'_y(x, y) \neq 0$ para $x \in (x_0 - \delta, x_0 + \delta)$, $y \in (y_0 - \varepsilon, y_0 + \varepsilon)$ con δ y ε suficientemente pequeños (¿por qué?), con lo que la fórmula (4.54) tiene sentido; se deriva también de ello que el teorema se puede aplicar en cualquier punto $(x, f(x))$, $x \in (x_0 - \delta, x_0 + \delta)$, lo que, usualmente, permitirá prolongar $y = f(x)$ a un intervalo más amplio que el original $(x_0 - \delta, x_0 + \delta)$.

Naturalmente, si $F'_y(x_0, y_0) = 0$ y $F'_x(x_0, y_0) \neq 0$ el teorema se aplica cambiando los papeles entre x e y para obtener una función implícita $x = g(y)$. La hipótesis clave es, pues, $\text{grad } F(x_0, y_0) \neq (0, 0)$. Bajo esta hipótesis, la gráfica de la ecuación $F(x, y) = 0$ en un entorno de (x_0, y_0) es la gráfica (una curva) de una función $y = f(x)$ o bien de una función $x = g(y)$, y estará justificado decir que $F(x, y) = 0$ define implícitamente una curva en un entorno de (x_0, y_0) .

En los ejemplos 1 y 2 se tiene $F'_y(x_0, y_0) = 2y_0 \neq 0$ para todo $y_0 \neq 0$, pero el teorema no es aplicable pues no hay punto $(x_0, y_0) \in \mathbb{R}^2$ en que se satisfagan las hipótesis: en el ejemplo 1 porque no hay ninguna solución de la ecuación y en el 2 porque en la única solución que existe no se verifica la hipótesis sobre la derivada.

En el ejemplo 3 se tiene $F'_y(x_0, y_0) = 3y_0^2 \neq 0$ para todo $y_0 \neq 0$; por ejemplo, $F(0, 1) = 0$, $F'_y(0, 1) = 3 \neq 0$, con lo que está definida $x \mapsto f(x)$ en un cierto intervalo $(-\delta, \delta)$; de hecho, en este caso, un cálculo algebraico elemental permite afirmar, como vimos, que hay una única función $y = f(x)$ tal que $F(x, f(x)) = 0$ que está definida para todo $x \in \mathbb{R}$ (también, por tanto, en las proximidades del punto $(1, 0)$, en el que no se verifica la hipótesis $F'_y(x_0, y_0) \neq 0$: las condiciones del teorema son suficientes pero no necesarias).

En el ejemplo 4, $F'_y(x_0, y_0) = 2y_0 \neq 0$ para todo (x_0, y_0) con $y_0 \neq 0$; en los puntos $(-1, 0)$ y $(1, 0)$, $F'_y = 0$ y, de hecho, no hay función $x \mapsto f(x)$ tal que $x^2 + f(x)^2 - 1 = 0$ en un intervalo $(\pm 1 - \delta, \pm 1 + \delta)$ con $\delta > 0$. Sin embargo, $F'_x(x_0, y_0) = 2x_0$ es distinta de cero en estos puntos y, cambiando los papeles entre x e y , el teorema de la función implícita garantiza la existencia de funciones $y \mapsto g(y)$ tales que $F(g(y), y) = 0$; se tiene, en efecto, $g(y) = (1 - y^2)^{1/2}$, $y \in (-1, 1)$, tal que $g(0) = 1$, y con el signo menos en la raíz alrededor del punto $(-1, 0)$.

En el ejemplo 5, $F_g(x, y) = x + 3y^2$, que es distinta de cero salvo en los puntos de la parábola $x + 3y^2 = 0$. A su vez, $F_x(x, y) = 3x^2 + y$. Se tiene $F_x = F_y = 0$ únicamente en los puntos $(0, 0)$ y $(-1/3, -1/3)$; el segundo no satisface la ecuación $F(x, y) = 0$, con lo que en cualquier punto (x_0, y_0) distinto del origen que satisfaga la ecuación está definida de modo único en un cierto intervalo o bien y en función de x o bien x en función de y . En la gráfica de la ecuación (realizada con un programa de ordenador. Fig. 4.5) se aprecia que en todo entorno del origen y queda definida como función de x de tres formas distintas (para cada $x < 0$ suficientemente pequeño existen tres raíces reales distintas de la ecuación en y de tercer grado $x^3 + xy + y^3 = 0$, mientras que para $x > 0$ sólo hay una raíz real y dos complejas conjugadas), y lo mismo ocurre para x como función de y .

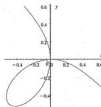


Figura 4.5

En el ejemplo 6 se tiene, por ejemplo,

$$F(0,1) = 0, \quad F_y(0,1) = 5y_0^4 - 2y_0 + 2x_0|_{(0,1)} = 3 \neq 0$$

con lo que existen $\delta > 0$, $\varepsilon > 0$ y una función $f: (-\delta, \delta) \rightarrow (1-\varepsilon, 1+\varepsilon)$ tales que $f(x)^5 - f(x)^2 + 2xf(x) - x^2 = 0$ para todo $x \in (-\delta, \delta)$; no se puede, sin embargo, expresar f en términos de funciones elementales. Sí es posible para esta ecuación expresar explícitamente x como función de y :

$$x = y \pm y^{\frac{5}{2}} \quad (4.55)$$

obteniéndose dos funciones definidas para $y \geq 0$. Obsérvese que se tiene

$$F(0,0) = F_x(0,0) = F_y(0,0) = 0$$

con lo que no se puede aplicar el teorema de la función implícita alrededor de la solución $(x_0, y_0) = (0,0)$; y, de hecho, aunque las hipótesis del teorema sólo son, como señalábamos, condiciones suficientes, se comprueba fácilmente en este caso, representando las curvas (4.55), que no es posible definir y como función continua $y = f(x)$ tal que $f(0) = 0$ ni x como función continua $x = g(y)$ tal que $g(0) = 0$ en intervalos centrados en 0.

En el ejemplo 7 se tiene, por ejemplo,

$$F(1,1) = 0, \quad F_y(1,1) = 1 - 2y|_{y=1} = -1 \neq 0$$

de modo que se tiene definida una función $y = f(x)$ en un cierto intervalo $(1-\delta, 1+\delta)$ tal que

$$x - 2 \ln x + f(x) - 2 \ln f(x) - 2 = 0, \quad x \in (1-\delta, 1+\delta)$$

pero no es posible expresarla explícitamente, ni tampoco a x como función explícita de y . Se puede demostrar que la gráfica de la ecuación (4.50) es una curva cerrada simple (Figura 4.6; véase [6]).



Figura 4.6

Obsérvese también que, una vez probado que la función $f(x)$ existe y tiene derivada continua, la fórmula (4.54) se obtiene aplicando la regla de la cadena a la función compuesta

$$x \mapsto \phi(x) = F(x, f(x)) = 0$$

En efecto, si introducimos provisionalmente una variable t que coincida con x , de forma que se tenga $x = t$, $y = f(t)$ (es decir, $u(t) = t$, $v(t) = f(t)$) se tendrá por el teorema 4.1

$$\phi'(t) = F_x(t, f(t)) \cdot 1 + F_y(t, f(t)) \cdot f'(t) = 0$$

de donde, despejando y volviendo a la variable x , se obtiene (4.54).

De la fórmula (4.54) se deriva otro hecho interesante; escrita en la forma

$$F'_x(x, f(x)) + F'_y(x, f(x))f'(x) = 0$$

o, equivalentemente,

$$(F'_x(x, f(x)), F'_y(x, f(x))) \cdot (1, f'(x)) = 0$$

nos indica que el gradiente de F es ortogonal a (la tangente a) la curva definida implícitamente por $F(x, y) = 0$ en cada uno de los puntos de ésta; basta recordar, en efecto, que $(1, f'(x))$ es un vector tangente a la curva $y = f(x)$ en el punto $(x, f(x))$. Esa propiedad geométrica se cumple también, naturalmente, si la ecuación define una función $x = g(y)$ en lugar de $y = f(x)$. Esto proporciona una interpretación geométrica muy elocuente de la hipótesis $F'_y(x_0, y_0) \neq 0$ del teorema: supuesto que $\text{grad} F(x_0, y_0) \neq (0, 0)$ (y que, por tanto, la gráfica de la ecuación $F(x, y) = 0$ en un entorno de (x_0, y_0) es una curva) la ecuación linealizada

$$\frac{\partial F}{\partial x}(x_0, y_0)(x - x_0) + \frac{\partial F}{\partial y}(x_0, y_0)(y - y_0) = 0$$

es la ecuación de la recta tangente en (x_0, y_0) a la curva definida implícitamente por $F(x, y) = 0$. La condición $F'_y(x_0, y_0) \neq 0$ implica que esa recta no es vertical, es decir, que es la gráfica de una función $y = l(x)$ y que sea razonable que esto mismo ocurra para la gráfica de $F(x, y) = 0$ (figura 4.7).

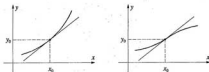


Figura 4.7

Por el contrario, si la recta tangente a la curva $F(x, y) = 0$ en (x_0, y_0) es vertical (o sea, $F'_y(x_0, y_0) = 0$), entonces esa curva puede o no ser la gráfica de una función $y = f(x)$, como se muestra en la figura 4.8 (la hipótesis $F'_y(x_0, y_0) \neq 0$ es suficiente pero no necesaria para que eso se dé). Las mismas observaciones valen para la condición $F'_x(x_0, y_0) \neq 0$ y una tangencia horizontal.



Figura 4.8

Curvas de nivel

Como sabemos, el conjunto de nivel c de la función $F(x, y)$ es el conjunto de soluciones de la ecuación $F(x, y) = c$, o sea

$$F(x, y) - c = 0$$

Si (x_0, y_0) es un punto del dominio (abierto) de F tal que

$$\text{grad } F(x_0, y_0) \neq (0, 0)$$

entonces el teorema de la función implícita implica que el conjunto de nivel de (x_0, y_0) , $F(x, y) - F(x_0, y_0) = 0$, es, en las proximidades de (x_0, y_0) , una

curva con derivada continua; y lo mismo ocurrirá para todo punto (x, y) de un entorno de (x_0, y_0) , siendo el vector $\text{grad } F(x, y)$, que señala la dirección de máximo crecimiento de la función F , ortogonal en cada punto a la curva de nivel que pasa por él (Fig. 4.9), según adelantábamos en el capítulo 3.



Figura 4.9

La tangente en el punto (x_0, y_0) a la curva de nivel de F que pasa por él es la correspondiente recta de nivel de la aproximación lineal

$$l(x, y) = F(x_0, y_0) + F'_x(x_0, y_0)(x - x_0) + F'_y(x_0, y_0)(y - y_0)$$

Si $F'_x(x_0, y_0) = F'_y(x_0, y_0) = 0$ se dice que (x_0, y_0) es un *punto crítico* de la función F , y no se puede asegurar que el correspondiente conjunto de nivel sea una curva. Por ejemplo, para la función

$$F(x, y) = x^2 + y^2$$

se tiene $F'_x(x, y) = 2x$, $F'_y(x, y) = 2y$; el único punto crítico es el origen $(0, 0)$ y el conjunto de nivel al que pertenece se reduce a ese punto. Por su parte, para la función

$$F(x, y) = y^3$$

se tiene $F'_x(x, y) = 0$, $F'_y(x, y) = 3y^2$ y los puntos críticos son los puntos $(x, 0)$ del eje x , el cual es el conjunto de nivel (cero) al que pertenecen todos ellos, que en este caso sí es una curva: $y = 0$ (las condiciones del teorema son suficientes pero no necesarias; considérense también las funciones $(x - y)^2$ y $(x - y)^3$ y sus conjuntos de nivel).

Conviene destacar algunos aspectos del teorema de la función implícita que han surgido en la discusión anterior:

- No hay que olvidar que para su aplicación hay que disponer de al menos una solución de la ecuación $F(x, y) = 0$: en la ecuación $x^2 + e^y = 0$ se tiene $F'_y = e^y \neq 0$ en todo punto del plano, pero no hay ningún punto (x_0, y_0) que satisfaga la ecuación y ésta no es resoluble en y .

- La condición $F'_y(x_0, y_0) \neq 0$ es suficiente pero no necesaria para la existencia de una función implícita $y = f(x)$ cuya gráfica pase por (x_0, y_0) , como hemos visto en el ejemplo 3; considérese también la ecuación $(x - y)^2 = 0$. De todos modos, si $F'_y = 0$ en todo un ε -entorno de (x_0, y_0) , entonces en ese entorno no hay definida ninguna función implícita $y = f(x)$ (¿por qué?; véase en el apartado anterior el comentario con que acaba el tema del teorema del valor medio).
- El teorema de la función implícita es un típico resultado de existencia de carácter local: afirma la existencia de una función implícita $y = f(x)$ pero, por sí mismo, no proporciona ninguna información sobre la forma concreta en que y depende de x ni da ninguna estimación sobre el tamaño del intervalo en que aquella está definida. De todos modos, como se apuntaba antes, sucesivas aplicaciones del teorema permiten con frecuencia extender la curva $y = f(x)$ a un intervalo más amplio y tener, en la práctica, un resultado global.

A pesar de estas observaciones, se trata de un resultado fundamental con multitud de implicaciones teóricas y prácticas. En el terreno de éstas últimas, nos permite, cuando se cumplen sus hipótesis, “trabajar” con la función $y = f(x)$ aunque en nuestro problema nos haya aparecido una ecuación $F(x, y) = 0$ en la que sea difícil o acaso imposible despejar y en términos de x . Podemos, por ejemplo, calcular su derivada mediante la fórmula (4.54). Así, para una función implícita $y = f(x)$ definida por ecuación

$$x^3 + xy + y^3 - 3 = 0 \quad (4.56)$$

se tiene

$$\frac{dy}{dx} = -\frac{3x^2 + y}{x + 3y^2} \quad (4.57)$$

cálculo válido en los puntos (x, y) tales que $x + 3y^2 \neq 0$ (los que verifican $x + 3y^2 = 0$ ya están excluidos de la aplicación del teorema). Desde luego, ocurrirá, en general, como en este caso, que la expresión para dy/dx contenga las dos variables x e y (pues, como hemos dicho, será frecuente que no se pueda despejar y como función explícita de x), pero para determinadas aplicaciones prácticas ello no será un inconveniente. Por ejemplo, podemos calcular la pendiente de la curva $y = f(x)$ en un punto de coordenadas (x_0, y_0) dadas (y si sólo se conoce una de ellas, no será difícil en muchos casos determinar la otra a partir de la ecuación). Así, para la curva $y = f(x)$ definida implícitamente por (4.56) que pasa por el punto $(1, 1)$ se tiene

$$\left. \frac{dy}{dx} \right|_{(1,1)} = -1$$

Obsérvese que en un entorno de dicho punto

$$\frac{dy}{dx} < 0$$

con lo que la curva es estrictamente decreciente en un entorno de $x_0 = 1$: recuérdese que en el estudio de una curva $y = f(x)$ lo que esencialmente utilizamos de la derivada es su signo, y no tanto su expresión explícita, para estudiar el crecimiento, puntos críticos,...; como veremos en el capítulo siguiente, se podrá, a su vez, derivar implícitamente en (4.57) para obtener una ecuación implícita en la derivada segunda y'' que, en situaciones favorables, permitirá un análisis de la concavidad de la función $y = f(x)$. En la figura 4.10 se muestra la gráfica de la ecuación (4.56).



Figura 4.10

NOTA. Para obtener la derivada de la función implícita no es necesario memorizar la fórmula (4.54). En la práctica, en cada caso concreto, los cálculos suelen hacerse en términos más informales que los utilizados antes para obtener dicha fórmula a partir del teorema 4.1. Por ejemplo, de

$$x^3 + xy + y^3 - 3 = 0$$

concluimos (haciendo mentalmente $y = f(x)$, derivando respecto a x y escribiendo $f' = y'$)

$$3x^2 + y + xy' + 3y^2y' = 0$$

de donde

$$y' = -\frac{3x^2 + y}{x + 3y^2}$$

De

$$x - 2\ln x + y - 2\ln y - 2 = 0$$

se obtiene

$$1 - \frac{2}{x} + y' - 2\frac{y'}{y} = 0$$

y de ahí

$$y' = -\frac{1 - \frac{2}{x}}{1 - \frac{2}{y}}$$

Nos proponemos ahora justificar mediante el teorema de la función implícita un hecho que, con seguridad, se habrá utilizado en cursos anteriores sobre bases meramente intuitivas, a saber, que F toma signos opuestos a uno y otro lado de la curva $F(x, y) = 0$. Por la unicidad que proporciona el teorema, en el rectángulo

$$R_{\delta, \varepsilon} = \{(x, y); x_0 - \delta < x < x_0 + \delta, y_0 - \varepsilon < y < y_0 + \varepsilon\}$$

los únicos puntos en los que $F = 0$, son los de la curva definida implícitamente por $F(x, y) = 0$; en los demás puntos, $F \neq 0$. Afirmamos que en $R_{\delta, \varepsilon}$ hay puntos en los que $F > 0$ y puntos en los que $F < 0$; de hecho, la curva divide al rectángulo en dos conjuntos disjuntos: en uno de ellos $F > 0$ y en el otro $F < 0$ (figura 4.11).

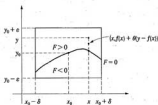


Figura 4.11

Supongamos, por ejemplo, que $F'_y > 0$ en $R_{\delta, \varepsilon}$. Sea $(x, y) \in R_{\delta, \varepsilon}$ tal que $y > f(x)$. Entonces, por el teorema del valor medio, existe $\theta \in (0, 1)$ tal que

$$F(x, y) = F(x, f(x)) + F'_x(x, f(x) + \theta(y - f(x))) \cdot 0 + F'_y(x, f(x) + \theta(y - f(x))) \cdot (y - f(x)) > 0$$

(en realidad, basta apelar al teorema del valor medio para funciones de una variable, la variable y).

Así pues, en todos los puntos de $R_{\delta, \varepsilon}$ que están por encima de la gráfica de $y = f(x)$ se tiene $F > 0$, mientras que, por un razonamiento análogo, se

tiene $F < 0$ en todos los que están por debajo (si fuese $F_y' < 0$, ocurriría lo contrario). Si la hipótesis es $F_x'(x_0, y_0) \neq 0$ se argumenta de igual manera según segmentos horizontales y F tendrá signos distintos a derecha e izquierda de la gráfica de la función implícita $x = g(y)$.

Podemos expresar el resultado anterior diciendo que, en $R_{\delta, \varepsilon}$, en los puntos que están a un lado de la curva, F tiene un signo determinado y en el otro lado, el signo opuesto, y esa expresión de "lado de la curva" tiene en $R_{\delta, \varepsilon}$, o sea, *localmente*, pleno sentido: si $F_y' \neq 0$, los puntos $\{(x, y) \in R_{\delta, \varepsilon}; y > f(x)\}$ constituyen uno de los lados y los puntos $\{(x, y) \in R_{\delta, \varepsilon}; y < f(x)\}$ el otro. Y lo análogo si es $F_x' \neq 0$. Diremos, en términos prácticos, que F cambia de signo al pasar por la curva $F(x, y) = 0$. Si no se verifican las hipótesis del teorema, no tiene por que ocurrir eso: la ecuación $(x - y)^2 = 0$ define implícitamente la función $y = x$ y a ambos lados de la recta $y = x$ la función $(x - y)^2$ es positiva (y, por otro lado, no se olvide que se trata de condiciones suficientes pero no necesarias para que se dé ese hecho: considérese la ecuación $(x - y)^3 = 0$).

En muchos problemas prácticos el conjunto de soluciones $\{(x, y); F(x, y) = 0\}$ es una curva en D en el entorno de cualquiera de sus puntos e interesa identificar en el plano el conjunto de puntos de D en los que $F(x, y) < 0$ y aquéllos en los que $F(x, y) > 0$. Pues bien, si en un punto $p \in D$ se tiene, por ejemplo, $F(p) > 0$, en cualquier otro punto q que se pueda unir con p mediante una poligonal contenida en D que no corte a la curva $F(x, y) = 0$ se tendrá también $F(q) > 0$; en efecto, si fuese $\text{signo } F(p) \neq \text{signo } F(q)$, habría dos vértices consecutivos de la poligonal, p_k, p_{k+1} , en los que F tendría signos opuestos; construyendo, como en la demostración del teorema del valor medio, la función de una variable

$$[0, 1] \ni t \longmapsto \phi(t) = F(p_k + t(p_{k+1} - p_k))$$

eso significa que $\text{signo } \phi(0) \neq \text{signo } \phi(1)$; pero $\phi(t)$ es continua por los resultados del capítulo 2 (es incluso derivable, según vimos en dicha demostración, y por tanto continua al tratarse de una función de una variable), por lo que existe $\theta \in (0, 1)$ tal que $\phi(\theta) = 0$; esto implica que tiene que haber un punto intermedio en el segmento $[p_k, p_{k+1}]$ en el que $F = 0$, contra la hipótesis de que la poligonal no corta a la curva. En la práctica, se elige un punto de coordenadas sencillas (el origen, usualmente, salvo que $F(0, 0) = 0$) y se comprueba el signo que tiene F en él; entonces, en todos los puntos que se puedan unir a él por una poligonal que no corte a $F(x, y) = 0$, F tendrá el mismo signo; por ejemplo, en el caso de la ecuación $x^2 + y^2 - 1 = 0$, la circunferencia que define divide al plano en dos regiones: la región en que $x^2 + y^2 - 1 < 0$, el círculo abierto delimitado por ella, y la región en que $x^2 + y^2 - 1 > 0$, el exterior de la circunferencia.

Funciones de más de dos variables

El teorema 4.11 se generaliza a funciones de más de dos variables:

Teorema 4.12 Sea $F(x_1, \dots, x_n, y)$ una función con derivadas parciales continuas en un conjunto abierto $D \subseteq \mathbb{R}^{n+1}$ y sea $(x_0, y_0) \equiv (x_1^0, \dots, x_n^0, y_0) \in D$ tal que $F(x_1^0, \dots, x_n^0, y_0) = 0$ y $F'_y(x_1^0, \dots, x_n^0, y_0) \neq 0$. Entonces existen $\delta > 0$, $\varepsilon > 0$ y una función $f: B_\delta(x_0) \rightarrow (y_0 - \varepsilon, y_0 + \varepsilon)$ tales que:

- i) $f(x_1^0, \dots, x_n^0) = y_0$
- ii) $F(x_1, \dots, x_n, f(x_1, \dots, x_n)) = 0$ para todo $x = (x_1, \dots, x_n) \in B_\delta(x_0)$
- iii) Para cada $x \in B_\delta(x_0)$, $y = f(x_1, \dots, x_n)$ es la única solución de la ecuación $F(x_1, \dots, x_n, y) = 0$ que pertenece a $(y_0 - \varepsilon, y_0 + \varepsilon)$.
- iv) La función $y = f(x_1, \dots, x_n)$ tiene derivadas parciales continuas en $B_\delta(x_0)$ que están dadas por

$$\frac{\partial f}{\partial x_i}(x_1, \dots, x_n) = -\frac{F'_{x_i}(x_1, \dots, x_n, y)}{F'_y(x_1, \dots, x_n, y)}, \quad i = 1, \dots, n \quad (4.58)$$

(En otras palabras, la ecuación $F(x_1, \dots, x_n, y) = 0$ define implícitamente a y como función de las variables x_1, \dots, x_n en un entorno de (x_1^0, \dots, x_n^0) , y ello de modo único).

La situación más frecuente en las aplicaciones será la de una ecuación en tres variables

$$F(x, y, z) = 0$$

que, bajo la hipótesis $F'_z(x_0, y_0, z_0) \neq 0$, define implícitamente una superficie $z = f(x, y)$ en el espacio \mathbb{R}^3 que pasa por (x_0, y_0, z_0) . En este caso, por (4.58), la ecuación del plano tangente en el punto (x_0, y_0, z_0) es

$$F'_x(x_0, y_0, z_0)(x - x_0) + F'_y(x_0, y_0, z_0)(y - y_0) + F'_z(x_0, y_0, z_0)(z - z_0) = 0$$

Vemos que el vector gradiente $\nabla F(x_0, y_0, z_0)$ es perpendicular al plano tangente, o sea, por definición, normal a la superficie $F(x, y, z) = 0$ en el punto (x_0, y_0, z_0) . (Fig. 4.12. Y, en general, análogamente al caso de dos variables, $\nabla F(x, y, z)$ es ortogonal a las superficies de nivel de la función de tres variables $F(x, y, z)$).

Si la hipersuperficie $y = f(x_1, \dots, x_n)$ está definida implícitamente por la ecuación $F(x_1, \dots, x_n, y) = 0$, su hiperplano tangente en el punto $(x_0, y_0) = (x_1^0, \dots, x_n^0, y_0)$ tiene por ecuación

$$F'_{x_1}(x_0, y_0)(x_1 - x_1^0) + \dots + F'_{x_n}(x_0, y_0)(x_n - x_n^0) + F'_y(x_0, y_0)(y - y_0) = 0$$

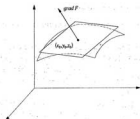


Figura 4.12

y, como para $n = 2$, el vector gradiente $\nabla F(x_1^0, \dots, x_n^0, y_0)$ es ortogonal a dicho hiperplano tangente.

Como antes, la fórmula (4.58) se deduce de la regla de la cadena aplicada a la función compuesta

$$(x_1, \dots, x_n) \mapsto G(x) = F(x_1, \dots, x_n, f(x_1, \dots, x_n, y)) \equiv 0$$

Por ejemplo, si $z = f(x, y)$ está definida implícitamente por la ecuación

$$F(x, y, z) = 3x^2z - x^2y^2 + 2z^3 + 3yz - 5 = 0$$

entonces

$$\begin{aligned} F'_x &= 6xz - 2xy^2 \\ F'_y &= -2x^2y + 3z \\ F'_z &= 3x^2 + 6z^2 + 3y \end{aligned}$$

y se tiene

$$\begin{aligned} \frac{\partial z}{\partial x} &= -\frac{F'_x}{F'_z} = -\frac{2xy^2 - 6xz}{3x^2 + 3y + 6z^2} \\ \frac{\partial z}{\partial y} &= -\frac{F'_y}{F'_z} = -\frac{2x^2y - 3z}{3x^2 + 3y + 6z^2} \end{aligned}$$

También como antes, estas derivadas parciales se calculan en la práctica derivando en la ecuación

$$3x^2z - x^2y^2 + 2z^3 + 3yz - 5 = 0$$

respecto a x y a y considerando mentalmente a z como función de las variables independientes x e y :

$$6xz + 3x^2z'_x - 2xy^2 + 6z^2z'_x + 3yz'_x = 0$$

$$3x^2z'_y - 2x^2y + 6z^2z'_y + 3z + 3yz'_y = 0$$

de donde, despejando,

$$z'_x = -\frac{6xz - 2xy^2}{3x^2 + 6z^2 + 3y}$$

$$z'_y = -\frac{-2x^2y + 3z}{3x^2 + 6z^2 + 3y}$$

Sistemas de ecuaciones

El teorema general de funciones implícitas se refiere a la posibilidad de resolver un sistema de m ecuaciones en $n + m$ variables

$$\begin{cases} F_1(x_1, \dots, x_n, y_1, \dots, y_m) = 0 \\ \vdots \\ F_m(x_1, \dots, x_n, y_1, \dots, y_m) = 0 \end{cases} \quad (4.59)$$

en las m variables y_1, \dots, y_m , es decir, a la cuestión de si este sistema de ecuaciones define implícitamente a las variables y_1, \dots, y_m como funciones de las variables x_1, \dots, x_n :

$$\begin{cases} y_1 = f_1(x_1, \dots, x_n) \\ \vdots \\ y_m = f_m(x_1, \dots, x_n) \end{cases}$$

Como decíamos antes, esta cuestión puede plantearse desde el punto de vista equivalente que consiste en preguntarse, supuesta conocida una solución $(x_0, y_0) = (x_1^0, \dots, x_n^0, y_1^0, \dots, y_m^0)$ del sistema de ecuaciones, si a cada $x = (x_1, \dots, x_n)$ próximo al punto $x_0 = (x_1^0, \dots, x_n^0)$ corresponde un $y = (y_1, \dots, y_m) = (f_1(x_1, \dots, x_n), \dots, f_m(x_1, \dots, x_n))$ tal que

$$(x, y) = (x_1, \dots, x_n, f_1(x_1, \dots, x_n), \dots, f_m(x_1, \dots, x_n))$$

sea también solución del sistema (4.59). Aplicando la misma idea que en el caso inicial de una ecuación en dos variables, podemos conjeturar que, si las funciones F_1, \dots, F_m son diferenciables, los resultados que se obtengan para el

sistema linealizado que resulta al sustituir F_1, \dots, F_m por sus aproximaciones lineales en (x_0, y_0) :

$$\begin{cases} \frac{\partial F_1}{\partial x_1}(x_0, y_0)(x_1 - x_1^0) + \dots + \frac{\partial F_1}{\partial x_n}(x_0, y_0)(x_n - x_n^0) + \\ + \frac{\partial F_1}{\partial y_1}(x_0, y_0)(y_1 - y_1^0) + \dots + \frac{\partial F_1}{\partial y_m}(x_0, y_0)(y_m - y_m^0) = 0 \\ \vdots \\ \frac{\partial F_m}{\partial x_1}(x_0, y_0)(x_1 - x_1^0) + \dots + \frac{\partial F_m}{\partial x_n}(x_0, y_0)(x_n - x_n^0) + \\ + \frac{\partial F_m}{\partial y_1}(x_0, y_0)(y_1 - y_1^0) + \dots + \frac{\partial F_m}{\partial y_m}(x_0, y_0)(y_m - y_m^0) = 0 \end{cases} \quad (4.60)$$

se podrán extender, al menos en las proximidades de (x_0, y_0) , al sistema (4.59). Como sabemos por álgebra lineal elemental, el sistema lineal (4.60) tiene solución única en las variables y_1, \dots, y_m para cada $x = (x_1, \dots, x_n)$ (y, en consecuencia, y_1, \dots, y_m quedan definidas explícitamente como funciones de x_1, \dots, x_n) si (y sólo si) es distinto de cero el determinante

$$\frac{\partial(F_1, \dots, F_m)}{\partial(y_1, \dots, y_m)} \stackrel{\text{def}}{=} \begin{vmatrix} \frac{\partial F_1}{\partial y_1}(x_0, y_0) & \dots & \frac{\partial F_1}{\partial y_m}(x_0, y_0) \\ \vdots & \dots & \vdots \\ \frac{\partial F_m}{\partial y_1}(x_0, y_0) & \dots & \frac{\partial F_m}{\partial y_m}(x_0, y_0) \end{vmatrix} = \det \left(\frac{\partial F_i}{\partial y_j}(x_0, y_0) \right) \quad (4.61)$$

llamado *determinante jacobiano* de las funciones F_1, \dots, F_m respecto a las variables y_1, \dots, y_m . Esta es, efectivamente, la hipótesis clave para demostrar, lo que haremos en el capítulo 9, el siguiente

Teorema 4.13 Sean $F_i(x_1, \dots, x_n, y_1, \dots, y_m)$, $i = 1, \dots, m$, m funciones con derivadas parciales continuas en un conjunto abierto $D \subseteq \mathbb{R}^{n+m}$ y sea $(x_0, y_0) \equiv (x_1^0, \dots, x_n^0, y_1^0, \dots, y_m^0) \in D$ tal que $F_i(x_1^0, \dots, x_n^0, y_1^0, \dots, y_m^0) = 0$ para todo $i = 1, \dots, m$ y

$$\det \left(\frac{\partial F_i}{\partial y_j}(x_0, y_0) \right)_{i,j=1,\dots,m} \neq 0$$

Entonces existen $\delta > 0$, $\varepsilon > 0$ y unas funciones de clase C^1 $f_j : B_\delta(x_0) \rightarrow (y_j^0 - \varepsilon, y_j^0 + \varepsilon)$, $j = 1, \dots, m$, tales que:

i)

$$\begin{cases} f_1(x_1^0, \dots, x_n^0) = y_1^0 \\ \vdots \\ f_m(x_1^0, \dots, x_n^0) = y_m^0 \end{cases}$$

ii)

$$\begin{cases} F_1(x_1, \dots, x_n, f_1(x_1, \dots, x_n), \dots, f_m(x_1, \dots, x_n)) = 0 \\ \vdots \\ F_m(x_1, \dots, x_n, f_1(x_1, \dots, x_n), \dots, f_m(x_1, \dots, x_n)) = 0 \end{cases} \quad (4.62)$$

para todo $x = (x_1, \dots, x_n) \in B_\delta(x_0)$ iii) Para cada $x = (x_1, \dots, x_n) \in B_\delta(x_0)$, $y = (y_1, \dots, y_m) = (f_1(x_1, \dots, x_n), \dots, f_m(x_1, \dots, x_n))$ es la única solución del sistema de ecuaciones

$$\begin{cases} F_1(x_1, \dots, x_n, y_1, \dots, y_m) = 0 \\ \vdots \\ F_m(x_1, \dots, x_n, y_1, \dots, y_m) = 0 \end{cases}$$

que pertenece a $\{y \in \mathbb{R}^m; |y_j - y_j^0| < \varepsilon\}$

(En otras palabras, el sistema de ecuaciones $F_i(x_1, \dots, x_n, y_1, \dots, y_m) = 0$ define implícitamente a las variables y_1, \dots, y_m como funciones de las variables x_1, \dots, x_n en un entorno de (x_1^0, \dots, x_n^0) , y ello de modo único).

Las derivadas de las funciones f_1, \dots, f_m , cuya existencia y continuidad garantiza el teorema, se obtienen, como en el caso de una sola ecuación, aplicando la regla de la cadena a las funciones compuestas

$$(x_1, \dots, x_n) \mapsto \phi_i(x_1, \dots, x_n) = F_i(x_1, \dots, x_n, f_1(x_1, \dots, x_n), \dots, f_m(x_1, \dots, x_n))$$

o sea, en términos prácticos, derivando implícitamente en las ecuaciones (4.62); así, para las derivadas respecto a x_1 se tiene:

$$\begin{cases} \frac{\partial F_1}{\partial x_1} + \frac{\partial F_1}{\partial y_1} \frac{\partial f_1}{\partial x_1} + \dots + \frac{\partial F_1}{\partial y_m} \frac{\partial f_m}{\partial x_1} = 0 \\ \vdots \\ \frac{\partial F_m}{\partial x_1} + \frac{\partial F_m}{\partial y_1} \frac{\partial f_1}{\partial x_1} + \dots + \frac{\partial F_m}{\partial y_m} \frac{\partial f_m}{\partial x_1} = 0 \end{cases} \quad (4.63)$$

que es un sistema lineal en $\left(\frac{\partial f_1}{\partial x_1}, \dots, \frac{\partial f_m}{\partial x_1}\right)$ cuyo determinante de coeficientes es precisamente el determinante jacobiano de F_1, \dots, F_m respecto a las variables y_1, \dots, y_m en puntos $(x, y) \in \{x \in B_\delta(x_0), |y_j - y_j^0| < \varepsilon\}$; dicho determinante es una función continua por ser suma de productos de derivadas primeras de F_1, \dots, F_m que, por hipótesis, son funciones continuas, por lo que, siendo, por hipótesis, distinto de cero en (x_0, y_0) tampoco se anula en puntos

próximos a (x_0, y_0) (reduciendo δ y ε si es necesario). Es posible, por tanto, obtener $\frac{\partial f_1}{\partial x_1}, \dots, \frac{\partial f_m}{\partial x_1}$ a partir de (4.63), y lo mismo para las derivadas respecto a x_2, \dots, x_n .

EJEMPLO 1. El sistema de ecuaciones

$$\begin{cases} F(x, y, z) \stackrel{\text{def}}{=} x + x^2 + y^3 - z = 0 \\ G(x, y, z) \stackrel{\text{def}}{=} xz^2 + y = 0 \end{cases}$$

define implícitamente a y y z como funciones de x : $y = f(x)$, $z = g(x)$ en las proximidades del punto $(0, 0, 0)$ ya que el determinante jacobiano

$$\frac{\partial(F, G)}{\partial(y, z)} = \begin{vmatrix} 3y^2 & -1 \\ 1 & 2xz \end{vmatrix} = 6xy^2z + 1$$

es distinto de cero en dicho punto. Para calcular las derivadas de f y g derivamos implícitamente en las ecuaciones dadas:

$$\begin{aligned} F_x + F_y \frac{dy}{dx} + F_z \frac{dz}{dx} &= 1 + 2x + 3y^2 \frac{dy}{dx} - \frac{dz}{dx} = 0 \\ G_x + G_y \frac{dy}{dx} + G_z \frac{dz}{dx} &= z^2 + \frac{dy}{dx} + 2xz \frac{dz}{dx} = 0 \end{aligned}$$

y obtenemos

$$\begin{aligned} \frac{dy}{dx} &= -\frac{\frac{\partial(F, G)}{\partial(x, z)}}{\frac{\partial(F, G)}{\partial(y, z)}} = -\frac{\begin{vmatrix} 1 + 2x & -1 \\ z^2 & 2xz \end{vmatrix}}{\begin{vmatrix} 3y^2 & -1 \\ 1 & 2xz \end{vmatrix}} = -z \frac{2x + 4x^2 + z}{6xy^2z + 1} \\ \frac{dz}{dx} &= -\frac{\frac{\partial(F, G)}{\partial(y, x)}}{\frac{\partial(F, G)}{\partial(y, z)}} = -\frac{\begin{vmatrix} 3y^2 & 1 + 2x \\ 1 & z^2 \end{vmatrix}}{\begin{vmatrix} 3y^2 & -1 \\ 1 & 2xz \end{vmatrix}} = \frac{1 + 2x - 3y^2z^2}{6xy^2z + 1} \end{aligned}$$

EJEMPLO 2. El sistema de ecuaciones

$$\begin{cases} F(x, y, u, v) \stackrel{\text{def}}{=} xv^2 + y^2u^3 - 1 = 0 \\ G(x, y, u, v) \stackrel{\text{def}}{=} 2xy^3 + u^2v = 0 \end{cases}$$

define implícitamente a u y v como funciones de x e y : $u = f(x, y)$, $v = g(x, y)$ en las proximidades del punto $(0, 1, 1, 0)$ ya que

$$\frac{\partial(F, G)}{\partial(u, v)}(0, 1, 1, 0) = \begin{vmatrix} 3y^2u^2 & 2xv \\ 2uv & u^2 \end{vmatrix}_{(0, 1, 1, 0)} = 3 \neq 0$$

Las derivadas parciales de f y g respecto a x se obtienen de

$$\begin{aligned} F_x + F_u \frac{\partial u}{\partial x} + F_v \frac{\partial v}{\partial x} &= v^2 + 3y^2 u^2 \frac{\partial u}{\partial x} + 2uv \frac{\partial v}{\partial x} = 0 \\ G_x + G_u \frac{\partial u}{\partial x} + G_v \frac{\partial v}{\partial x} &= 2y^3 + 2uv \frac{\partial u}{\partial x} + u^2 \frac{\partial v}{\partial x} = 0 \end{aligned}$$

de donde

$$\frac{\partial u}{\partial x} = -\frac{\frac{\partial(F,G)}{\partial(x,v)}}{\frac{\partial(F,G)}{\partial(u,v)}} = -\frac{\begin{vmatrix} v^2 & 2uv \\ 2y^3 & u^2 \end{vmatrix}}{\begin{vmatrix} 3y^2 u^2 & 2uv \\ 2uv & u^2 \end{vmatrix}}, \quad \frac{\partial v}{\partial x} = -\frac{\frac{\partial(F,G)}{\partial(x,u)}}{\frac{\partial(F,G)}{\partial(u,v)}} = -\frac{\begin{vmatrix} 3y^2 u^2 & v^2 \\ 2uv & 2y^3 \end{vmatrix}}{\begin{vmatrix} 3y^2 u^2 & 2uv \\ 2uv & u^2 \end{vmatrix}}$$

De manera análoga

$$\frac{\partial u}{\partial y} = -\frac{\frac{\partial(F,G)}{\partial(y,v)}}{\frac{\partial(F,G)}{\partial(u,v)}} = -\frac{\begin{vmatrix} 2yu^3 & 2uv \\ 6xy^2 & u^2 \end{vmatrix}}{\begin{vmatrix} 3y^2 u^2 & 2uv \\ 2uv & u^2 \end{vmatrix}}, \quad \frac{\partial v}{\partial y} = -\frac{\frac{\partial(F,G)}{\partial(y,u)}}{\frac{\partial(F,G)}{\partial(u,v)}} = -\frac{\begin{vmatrix} 3y^2 u^2 & 2yu^3 \\ 2uv & 6xy^2 \end{vmatrix}}{\begin{vmatrix} 3y^2 u^2 & 2uv \\ 2uv & u^2 \end{vmatrix}}$$

El teorema de la función inversa

Como decíamos en el primer apartado, es frecuente introducir en el estudio de determinadas funciones un *cambio de variables*

$$\begin{cases} x = u(s, t) \\ y = v(s, t) \end{cases} \quad (4.64)$$

Para que dicho cambio sea realmente útil, es necesario (como se recordará de cursos anteriores: cálculo de primitivas, cambios de base en álgebra lineal,...) que la correspondencia $(x, y) \leftrightarrow (s, t)$ sea *inyectiva* (o uno a uno), en otras palabras, que la aplicación de \mathbb{R}^2 en \mathbb{R}^2 dada por (4.64) sea *invertible*, que existan dos funciones

$$\begin{cases} s = f(x, y) \\ t = g(x, y) \end{cases}$$

de modo que

$$\begin{cases} x = u(f(x, y), g(x, y)) \\ y = v(f(x, y), g(x, y)) \end{cases} \quad \begin{cases} s = f(u(s, t), v(s, t)) \\ t = g(u(s, t), v(s, t)) \end{cases}$$

En general, si se tienen n funciones

$$\begin{cases} y_1 = f_1(x_1, \dots, x_n) \\ \vdots \\ y_n = f_n(x_1, \dots, x_n) \end{cases} \quad (4.65)$$

se trata de estudiar si a cada (y_1, \dots, y_n) le corresponde un $(x_1, \dots, x_n) = (g_1(y_1, \dots, y_n), \dots, g_n(y_1, \dots, y_n))$ unívocamente determinado de modo que

$$\begin{cases} x_1 = g_1(f_1(x_1, \dots, x_n), \dots, f_n(x_1, \dots, x_n)) \\ \vdots \\ x_n = g_n(f_1(x_1, \dots, x_n), \dots, f_n(x_1, \dots, x_n)) \\ y_1 = f_1(g_1(y_1, \dots, y_n), \dots, g_n(y_1, \dots, y_n)) \\ \vdots \\ y_n = f_n(g_1(y_1, \dots, y_n), \dots, g_n(y_1, \dots, y_n)) \end{cases} \quad (4.66)$$

Este es el problema análogo al de invertir una función de una variable $y = f(x)$, ahora relativo a la función vectorial $f = (f_1, \dots, f_n)$ de n variables (x_1, \dots, x_n) . Para funciones de una variable $f: I \rightarrow \mathbb{R}$, I un intervalo de \mathbb{R} , la condición de monotonía estricta (creciente o decreciente) es claramente una condición suficiente para la existencia de la función inversa f^{-1} (si f es continua en I , la condición es también necesaria). Si la función f es derivable, esa condición se comprueba normalmente analizando el signo de f' en I , y, así, se tiene que si $f' \neq 0$ en I (lo que implica que f' es siempre positiva o siempre negativa en I , aunque no se suponga que f' sea continua), entonces f es estrictamente monótona en I y la función inversa f^{-1} es derivable en todo punto $y = f(x) \in f(I)$ verificándose que

$$(f^{-1})'(y) = \frac{1}{f'(x)}$$

Para funciones vectoriales se tiene un resultado semejante contemplando el problema como el de la resolución única en x_1, \dots, x_n , para cada y_1, \dots, y_n , del sistema de ecuaciones

$$\begin{cases} f_1(x_1, \dots, x_n) - y_1 = 0 \\ \vdots \\ f_n(x_1, \dots, x_n) - y_n = 0 \end{cases}$$

y aplicando el teorema de funciones implícitas:

Teorema 4.14 (Teorema de la función inversa) Sean $f_1, \dots, f_n: \Omega \rightarrow \mathbb{R}^n$ n funciones de clase $C^1(\Omega)$, Ω un abierto de \mathbb{R}^n y sea $x_0 = (x_1^0, \dots, x_n^0) \in \Omega$ tal que el determinante de la matriz

$$Df(x_0) \stackrel{\text{def}}{=} \left(\frac{\partial f_i}{\partial x_j}(x_0) \right), \quad i, j = 1, \dots, n$$

es distinto de cero. Entonces existe un $\delta > 0$ tal que en la bola $B_\delta(y_0)$, donde $y_0 = f(x_0) = (f_1(x_0), \dots, f_n(x_0))$, están definidas unas funciones g_1, \dots, g_n de clase C^1 que verifican (4.66) para cada $(y_1, \dots, y_n) \in B_\delta(y_0)$. En particular, $(g_1(y_0), \dots, g_n(y_0)) = x_0$.

La función vectorial $g = (g_1, \dots, g_n)$ es la inversa en $B_\delta(y_0)$ de la función $f = (f_1, \dots, f_n)$ y se representa por f^{-1} . La matriz

$$Df(x_0) \stackrel{\text{def}}{=} \left(\frac{\partial f_i}{\partial x_j}(x_0) \right), \quad i, j = 1, \dots, n$$

es la matriz jacobiana de f en x_0 y su determinante, el determinante jacobiano en x_0 , se representa por $Jf(x_0)$ o, como ya dijimos, por $\frac{\partial(f_1, \dots, f_n)}{\partial(x_1, \dots, x_n)}$ o $\frac{\partial(y_1, \dots, y_n)}{\partial(x_1, \dots, x_n)}$. Las derivadas de la función inversa $g(y)$ se calculan por derivación implícita en (4.65), teniéndose para cada $y = f(x) \in B_\delta(y_0)$ la siguiente relación entre matrices jacobianas

$$Dg(y) = [Df(x)]^{-1}$$

NOTAS. 1. También aquí cabe señalar, como hacíamos a propósito del teorema de la función implícita, que el de la función inversa es un teorema de existencia: garantiza la existencia de la función inversa pero no proporciona una fórmula explícita para ella. Por ejemplo, la función de una variable

$$f(x) = e^x (3 + x^2 - 2x)$$

verifica $f'(x) = (1 + x^2) e^x > 0$ para todo $x \in \mathbb{R}$ y, por tanto, es estrictamente creciente en todo \mathbb{R} y, por ello, inyectiva; además, $f(x) \rightarrow \infty$ cuando $x \rightarrow \infty$ y $f(x) \rightarrow 0$ cuando $x \rightarrow -\infty$. Consecuentemente, f aplica \mathbb{R} sobre $(0, \infty)$ y está bien definida $f^{-1}: (0, \infty) \rightarrow \mathbb{R}$ pero no puede darse para ella una fórmula explícita en términos de funciones elementales.

2. Bajo las hipótesis del teorema de la función implícita, la aplicación que lleva de x a $f(x)$ es invertible localmente, en un entorno —acaso pequeño— del punto x_0 en el que $Jf(x_0) \neq 0$, pero, en general, no lo será globalmente, o sea, en todo el dominio de la función f : en dimensiones superiores a 1 no cabe esperar obtener resultados globales tan fácilmente como en el ejemplo unidimensional anterior. Sea, por ejemplo, la función

$$f(x_1, x_2) = \begin{pmatrix} f_1(x_1, x_2) \\ f_2(x_1, x_2) \end{pmatrix} = \begin{pmatrix} e^{x_1} \cos x_2 \\ e^{x_1} \sin x_2 \end{pmatrix}$$

$$\det Df(x_1, x_2) = \begin{vmatrix} e^{x_1} \cos x_2 & -e^{x_1} \sin x_2 \\ e^{x_1} \sin x_2 & e^{x_1} \cos x_2 \end{vmatrix}$$

es distinto de cero para todo $(x_1, x_2) \in \mathbb{R}^2$ con lo que $f(x_1, x_2)$ es inyectiva en las proximidades de cualquier punto del plano. Sin embargo, no es inyectiva sobre el plano \mathbb{R}^2 entero pues

$$f(x_1, x_2) = f(x_1, x_2 + 2\pi).$$

EJEMPLO 3. Estudiemos si es invertible el sistema

$$\begin{aligned} y_1 &= x_1 - 2x_2 + x_1x_2 \\ y_2 &= x_2 + x_1^2 + \sin(x_1x_2) \end{aligned} \quad (4.67)$$

en un entorno de $(0, 0, 0, 0)$. Se tiene

$$\begin{aligned} \frac{\partial(y_1, y_2)}{\partial(x_1, x_2)}(0, 0) &= \begin{vmatrix} 1 + x_2 & -2 + x_1 \\ 2x_1 + (\cos x_1x_2)x_2 & 1 + (\cos x_1x_2)x_1 \end{vmatrix}_{(0,0)} = \\ &= \begin{vmatrix} 1 & -2 \\ 0 & 1 \end{vmatrix} \neq 0 \end{aligned}$$

con lo que, efectivamente, el sistema tiene solución única $x_1 = g_1(y_1, y_2)$, $x_2 = g_2(y_1, y_2)$ para cada (y_1, y_2) suficientemente próximo a $(0, 0)$ y las funciones así obtenidas satisfacen $g_1(0, 0) = 0$, $g_2(0, 0) = 0$; en otras palabras, la función vectorial $f = (f_1, f_2)$ tiene inversa en un entorno de $f(0, 0) = (f_1(0, 0), f_2(0, 0)) = (0, 0)$. Para obtener las derivadas de g_1 y g_2 , derivamos implícitamente en (4.67) respecto a y_1 e y_2 :

$$\begin{aligned} 1 &= \frac{\partial g_1}{\partial y_1} - 2 \frac{\partial g_2}{\partial y_1} + x_2 \frac{\partial g_1}{\partial y_1} + x_1 \frac{\partial g_2}{\partial y_1} \\ 0 &= \frac{\partial g_2}{\partial y_1} + 2x_1 \frac{\partial g_1}{\partial y_1} + (\cos x_1x_2)x_2 \frac{\partial g_1}{\partial y_1} + (\cos x_1x_2)x_1 \frac{\partial g_2}{\partial y_1} \end{aligned}$$

o sea

$$\begin{aligned} (1 + x_2) \frac{\partial g_1}{\partial y_1} + (-2 + x_1) \frac{\partial g_2}{\partial y_1} &= 1 \\ (2x_1 + (\cos x_1x_2)x_2) \frac{\partial g_1}{\partial y_1} + (1 + (\cos x_1x_2)x_1) \frac{\partial g_2}{\partial y_1} &= 0 \end{aligned}$$

de donde

$$\frac{\partial g_1}{\partial y_1} = \frac{\begin{vmatrix} 1 & -2 + x_1 \\ 0 & 1 + (\cos x_1x_2)x_1 \end{vmatrix}}{\begin{vmatrix} 1 + x_2 & -2 + x_1 \\ 2x_1 + (\cos x_1x_2)x_2 & 1 + (\cos x_1x_2)x_1 \end{vmatrix}} = \frac{1 + (\cos x_1x_2)x_1}{Jf(x_1, x_2)}$$

y

$$\frac{\partial g_2}{\partial y_1} = \frac{\begin{vmatrix} 1 + x_2 & 1 \\ 2x_1 + (\cos x_1x_2)x_2 & 0 \end{vmatrix}}{\begin{vmatrix} 1 + x_2 & -2 + x_1 \\ 2x_1 + (\cos x_1x_2)x_2 & 1 + (\cos x_1x_2)x_1 \end{vmatrix}} = \frac{2x_1 + (\cos x_1x_2)x_2}{Jf(x_1, x_2)}$$

y análogamente para las derivadas respecto a y_2 . En general, naturalmente, se obtendrán expresiones implícitas para las derivadas ya que $x_1 = g_1(y_1, y_2)$, $x_2 = g_2(y_1, y_2)$ no se conocerán explícitamente. Compruébese en este caso que las matrices $Df(x_1, x_2)$ y $Dg(f_1(x_1, x_2), f_2(x_1, x_2))$ son inversas una de la otra.

Dependencia funcional

Cuando se emplea la idea de función en las aplicaciones prácticas, se usan frecuentemente expresiones como "la variable y depende funcionalmente de la variable x " para indicar la existencia de una función f (en el sentido formal del término) tal que $y = f(x)$ (o $x = f(y)$). En este contexto, x e y son magnitudes reales que describen ciertos aspectos del fenómeno que estamos estudiando, y f es la "fórmula" que permite obtener una de ellas una vez conocido el valor de la otra. Naturalmente, en el tratamiento general, tal "fórmula" no tiene por qué existir (recuérdese el problema suscitado en relación a las funciones definidas implícitamente por ecuaciones como $F(x, y) = 0$, cuya existencia se puede demostrar bajo ciertas hipótesis pero que son, en general, imposibles de expresar en términos de las funciones habituales), y por eso la noción moderna de función, se conforma con exigir la *univocidad*, es decir, el carácter único del y asociado a cada x , sea éste calculable mediante una fórmula o no.

Así, en la descripción cuantitativa de los fenómenos reales, aparece una multitud de variables asociadas a las diversas magnitudes que, por un proceso de abstracción, permiten descomponer el proceso complejo en elementos más simples, de forma que su modelización matemática podría describirse como el estudio de las diversas *relaciones funcionales* que se dan entre esas variables, o al menos entre las más relevantes.

Ahora bien, este proceso de obtención de relaciones funcionales no tiene por qué quedarse al primer nivel: no necesariamente existen unas variables que puedan considerarse como *independientes* en términos absolutos, dependiendo las demás de ellas: pueden existir niveles superpuestos descritos mediante *funciones compuestas* del tipo estudiado en el apartado 4.1.

Por ejemplo, en la *teoría cinética de los gases*, las variables llamadas *macroscópicas* (presión (p), volumen (V) y temperatura (T), entre otras) se explican en términos de las velocidades de las moléculas que componen el gas y de los choques entre ellas, y vienen definidas matemáticamente en términos de *cuatrillones* de variables, que son las posiciones y velocidades de todas esas moléculas. Si llamamos $x = (x_1, x_2, \dots)$ al vector (de dimensión del orden de un cuatrillón: 10^{24}) que reúne todas esas variables, la mencionada teoría proporciona fórmulas f_1, f_2, f_3 que permiten expresar T, p y V en función de x : $p = f_1(x), V = f_2(x), T = f_3(x)$. Pues bien, existe una constante conocida k tal que *siempre se verifica que* $T = kpV$ *con independencia de las posiciones y velocidades* x . Es decir, $f_3(x) = kf_1(x)f_2(x)$ para todo x . Esta es una relación funcional *macroscópica* que no depende del conocimiento (imposible, por otra parte) de las posiciones y velocidades de las más de 10^{23} moléculas del gas. Es, pues, difícil de exagerar la importancia de este tipo de relaciones "a nivel superior" para la adecuada comprensión del fenómeno.

También en el ámbito de la economía se dan relaciones entre variables agre-

gadas, obtenidas a partir de las interacciones, deseos e intereses de millones de consumidores y productores que, sin embargo, parecen ponerse de acuerdo en generar tanto una "demanda global" o "agregada" de un determinado bien de consumo como la correspondiente "oferta global" del mismo. Las relaciones que, con independencia del sustrato individual que las genera, se puedan establecer entre estas variables globales serán de la mayor importancia en el estudio del mercado del bien de consumo en cuestión.

Indiquemos finalmente que la dependencia funcional de la que aquí hablamos puede entenderse como un caso extremo de la *dependencia estadística* que se estudia, por ejemplo, en el cálculo de *rectas de regresión*: dadas unas variables estadísticas x e y , cuya representación sobre un plano xy genera una nube de puntos, se trata de ver qué fórmula de primer grado $y = ax + b$ es la que *mejor se ajusta* a los datos: o sea, qué recta representa mejor a la nube de puntos (véase el apartado 6.1). Desde nuestro punto de vista, para decir que x e y se hallan en la relación funcional $y = ax + b$ exigiríamos que *todos* los puntos de la nube se hallasen *exactamente* sobre la recta $y = ax + b$.

De hecho, esta comparación con el análisis estadístico nos puede arrojar mucha luz sobre el problema que nos ocupa. En efecto, en el caso de la teoría cinética de gases, la relación macroscópica $T = kpV$ se puede interpretar así: Dibújense, en un sistema de tres ejes (p, V, T) , las ternas de valores de presión, volumen y temperatura correspondientes a *todas* las posibles posiciones y velocidades de las moléculas del gas. Es decir, dibújese el conjunto de las ternas $(f_1(x), f_2(x), f_3(x))$ para *todos* los vectores x de posiciones y velocidades. La relación de dependencia funcional *exacta* (en oposición a *estadística*) implica que *todos* esos puntos se encuentran *sobre la superficie* $T = kpV$ (en forma de hiperboloide: véase el capítulo 1).

Veamos un ejemplo sencillo: Sean las funciones $u = f(x, y) \stackrel{\text{def}}{=} x + y$, $v = g(x, y) \stackrel{\text{def}}{=} x^2 + 2xy + y^2$.

Entonces: $g(x, y) = f(x, y)^2$ para *toda* (x, y) , lo cual abreviamos diciendo que $v = u^2$. Geométricamente, si dibujamos en un plano (u, v) los puntos de la forma $(x + y, x^2 + 2xy + y^2)$, todos ellos estarán sobre la parábola $v = u^2$. Decimos en este caso que v depende funcionalmente de u (o que v "es función de u ", pues el conocimiento de $u = x + y$ es suficiente para obtener v *sin necesidad* de conocer (x, y)).

En cambio, si $w = h(x, y) \stackrel{\text{def}}{=} x - y$, vemos que, a medida que vamos dando valores a (x, y) , los correspondientes puntos $(x + y, x - y)$ del plano (u, w) no se quedan en una curva de dicho plano, sino que lo van llenando por todas partes, reflejo de la *independencia funcional* de u y w . Desde el punto de vista aritmético, lo que esto nos dice es que el conocimiento de la suma de dos números no nos da pista alguna para obtener su diferencia:

$7 + 3 = 10$ y $7 - 3 = 4$, mientras que $3 + 7 = 10$ y $3 - 7 = -4$. En cuanto tengamos dos pares (x_1, y_1) , (x_2, y_2) para los cuales $f(x_1, y_1) = f(x_2, y_2)$ pero $h(x_1, y_1) \neq h(x_2, y_2)$, ya sabemos que una dependencia funcional del tipo $w = \varphi(u)$ (o sea, $h(x, y) = \varphi(f(x, y))$) para *todo* (x, y) es imposible, pues se tendría: $h(x_2, y_2) = \varphi(f(x_2, y_2)) = \varphi(f(x_1, y_1)) = h(x_1, y_1)$, contra la hipótesis de que $h(x_1, y_1) \neq h(x_2, y_2)$.

En general, si f y g son dos funciones de clase C^1 en un abierto $\Omega \subseteq \mathbb{R}^2$, se dice que g depende funcionalmente de f si existe una función de clase C^1 tal que

$$g(x, y) = \varphi(f(x, y)) \text{ para todo } (x, y) \in \Omega \quad (4.68)$$

Si derivamos en (4.68), obtenemos por la regla de la cadena

$$\begin{cases} \varphi' \frac{\partial f}{\partial x} - \frac{\partial g}{\partial x} = 0 \\ \varphi' \frac{\partial f}{\partial y} - \frac{\partial g}{\partial y} = 0 \end{cases} \quad (4.69)$$

En cada $(x, y) \in \Omega$, el sistema lineal homogéneo (4.69) tiene la solución no trivial $(\varphi'(f(x, y)), -1)$, por lo que el determinante de los coeficientes ha de ser nulo:

$$\det \begin{pmatrix} \frac{\partial f}{\partial x} & \frac{\partial g}{\partial x} \\ \frac{\partial f}{\partial y} & \frac{\partial g}{\partial y} \end{pmatrix} = 0$$

o, equivalentemente,

$$\det \begin{pmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} \\ \frac{\partial g}{\partial x} & \frac{\partial g}{\partial y} \end{pmatrix} = 0 \quad (4.70)$$

(En el ejemplo anterior:

$$\det \begin{pmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} \\ \frac{\partial g}{\partial x} & \frac{\partial g}{\partial y} \end{pmatrix} = \begin{vmatrix} 1 & 1 \\ 2x+2y & 2x+2y \end{vmatrix} = (2x+2y) - (2x+2y) = 0$$

para todo $(x, y) \in \mathbb{R}^2$.)

A la misma condición (4.70) se llega si suponemos que la función f depende de g : $f(x, y) = \psi(g(x, y))$ (en el ejemplo, $f(x, y) = \sqrt{g(x, y)}$ o $f(x, y) = -\sqrt{g(x, y)}$). Diremos que dos funciones f, g de clase $C^1(\Omega)$ son *funcionalmente dependientes* en Ω si una de ellas depende funcionalmente de la otra. En caso contrario, se dice que son *funcionalmente independientes*. Hemos comprobado que para que dos funciones f y g sean funcionalmente dependientes en Ω es condición necesaria el que se anule en todo $(x, y) \in \Omega$ el jacobiano de f y g respecto a x e y :

$$\frac{\partial(f, g)}{\partial(x, y)} = \begin{vmatrix} \frac{\partial f}{\partial x}(x, y) & \frac{\partial f}{\partial y}(x, y) \\ \frac{\partial g}{\partial x}(x, y) & \frac{\partial g}{\partial y}(x, y) \end{vmatrix} = 0 \quad (4.71)$$

Como veremos a continuación, esta condición es también “casi” suficiente; es una condición muy fácil de verificar en la práctica y esta es la razón para suponer la regularidad C^1 en la definición de dependencia funcional.

Concretamente se tiene que si se verifica (4.71) para todo $(x, y) \in \Omega$, siendo Ω un abierto conexo de \mathbb{R}^2 , entonces f y g son funcionalmente dependientes en un cierto subconjunto abierto de Ω . Veamos. Si todos los elementos de $\frac{\partial(f, g)}{\partial(x, y)}$, o sea, todas las derivadas de f y g , se anulan en todo punto $(x, y) \in \Omega$, entonces f y g son constantes en Ω (la hipótesis de conexión se hace para obtener esto) y hay infinitas formas de expresar $g(x, y) = \varphi(f(x, y))$ (o $f(x, y) = \psi(g(x, y))$) válidas, de hecho, para todo $(x, y) \in \Omega$.

Supongamos entonces que una al menos de las derivadas

$$\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial g}{\partial x}, \frac{\partial g}{\partial y}$$

no es idénticamente nula en Ω . Sea, por ejemplo,

$$\frac{\partial f}{\partial y}(x_0, y_0) \neq 0 \quad (4.72)$$

(en otro caso, los argumentos serían totalmente análogos) para un cierto punto $(x_0, y_0) \in \Omega$ y sea $u_0 = f(x_0, y_0)$.

Por el teorema de la función implícita, existen $\delta > 0$, $\varepsilon > 0$ tales que la ecuación en tres variables

$$f(x, y) = u$$

define, para (x, u) tales que $|x - x_0| < \delta$, $|u - u_0| < \varepsilon$, una única función de clase C^1

$$y = \phi(x, u) \quad (4.73)$$

que toma valores en el intervalo $(y_0 - \varepsilon, y_0 + \varepsilon)$. Entonces, para (x, y) tales que $|x - x_0| < \delta$, $|y - y_0| < \varepsilon$, se tiene

$$g(x, y) = g(x, \phi(x, u)) \quad (4.74)$$

Si probamos que $G(x, u) = g(x, \phi(x, u))$ es independiente de x , o sea, que es de la forma $\varphi(u)$ en el cuadrado $\{(x, y); |x - x_0| < \delta, |y - y_0| < \varepsilon\}$, entonces se tendrá

$$g(x, y) = \varphi(f(x, y)) \quad (4.75)$$

(pues los valores u tales que $|u - u_0| < \delta$ son exactamente los valores que toma $f(x, y)$ cuando $(x, y) \in \{(x, y) \in \Omega; |x - x_0| < \delta, |y - y_0| < \varepsilon\}$). Pero (4.75) quiere decir que f y g son funcionalmente dependientes en el abierto $\{(x, y) \in \Omega; |x - x_0| < \delta, |y - y_0| < \varepsilon\}$, que es lo que deseábamos obtener.

Por la regla de la cadena aplicada a (4.74):

$$\frac{\partial G}{\partial x} = \frac{\partial g}{\partial x} + \frac{\partial g}{\partial y} \frac{\partial \phi}{\partial x} \quad (4.76)$$

Pero, aplicando también la regla de la cadena a la función compuesta

$$(x, u) \mapsto f(x, \phi(x, u)) - u = 0 \quad (4.77)$$

(como en ocasiones anteriores) se obtiene

$$\frac{\partial f}{\partial x} + \frac{\partial f}{\partial y} \frac{\partial \phi}{\partial x} = 0 \quad (4.78)$$

Utilizamos ahora la hipótesis de que el jacobiano se anula

$$\begin{vmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} \\ \frac{\partial g}{\partial x} & \frac{\partial g}{\partial y} \end{vmatrix} = 0$$

lo que significa que existe $\lambda (= \lambda(x, y))$ tal que

$$\frac{\partial g}{\partial x} = \lambda \frac{\partial f}{\partial x}, \quad \frac{\partial g}{\partial y} = \lambda \frac{\partial f}{\partial y}$$

(recuérdese que $\partial f / \partial y \neq 0$ en un entorno de (x_0, y_0) , por la continuidad de $\partial f / \partial y$) y, en consecuencia, por (4.76) y (4.78), que

$$\frac{\partial G}{\partial x} = \lambda \left(\frac{\partial f}{\partial x} + \frac{\partial f}{\partial y} \frac{\partial \phi}{\partial x} \right) = 0$$

Pero esto implica (véanse los comentarios que siguen al teorema del valor medio) que $G(x, u) = g(x, \phi(x, u))$ es independiente de x en

$$\{(x, u); |x - x_0| < \delta, |u - u_0| < \delta\}$$

Con los mismos argumentos, podemos también establecer una condición suficiente para la dependencia funcional local, o sea en el entorno de cada punto de Ω :

Teorema 4.15 Sean f y g dos funciones de clase C^1 en un abierto Ω de \mathbb{R}^2 . Si

$$\text{rango} \begin{pmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} \\ \frac{\partial g}{\partial x} & \frac{\partial g}{\partial y} \end{pmatrix} = 1 \quad (4.79)$$

para todo $(x, y) \in \Omega$, entonces f y g son funcionalmente dependientes en un cierto entorno de cada punto $(x, y) \in \Omega$.

En efecto, considerando un $(x_0, y_0) \in \Omega$ cualquiera, la hipótesis (4.79) implica que al menos una de las derivadas

$$\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial g}{\partial x}, \frac{\partial g}{\partial y}$$

es distinta de cero en el punto (x_0, y_0) . Ello permite, como antes, encontrar una relación funcional entre f y g en un conjunto abierto $\{(x, y) \in \Omega; |x - x_0| < \delta, |y - y_0| < \varepsilon\}$ que contiene a (x_0, y_0) .

EJEMPLO 4. Analicemos si existe dependencia funcional entre las funciones

$$f(x, y) = 2xy + 2x + 1, \quad g(x, y) = x^2y^2 + 2x^2y + x^2 - 1$$

Se tiene

$$\begin{aligned} \frac{\partial(f, g)}{\partial(x, y)} &= \begin{vmatrix} 2y+2 & 2x \\ 2xy^2+4xy+2x & 2x^2y+2x^2 \end{vmatrix} = \\ &= \begin{vmatrix} 2(y+1) & 2x \\ 2x(y+1)^2 & 2x^2(y+1) \end{vmatrix} = 0 \end{aligned}$$

para todo $(x, y) \in \mathbb{R}^2$. Observamos que el rango es 1 en todo punto del conjunto abierto $\mathbb{R}^2 \setminus \{(0, -1)\}$, por lo que f y g son funcionalmente dependientes

en un entorno de cada uno de esos puntos. Intentemos obtener la relación que las liga siguiendo la pauta de la demostración anterior. De

$$2xy + 2x + 1 = u \Leftrightarrow 2x(y + 1) = u - 1$$

resulta (para $(x, y) \neq (0, -1)$)

$$y = \frac{u - 1}{2x} - 1$$

que sustituido en

$$v = x^2 y^2 + 2x^2 y + x^2 - 1$$

arroja

$$v = x^2 \left(\frac{u - 1}{2x} - 1 \right)^2 + 2x^2 \left(\frac{u - 1}{2x} - 1 \right) + x^2 - 1 = \frac{1}{4}u^2 - \frac{1}{2}u - \frac{3}{4}$$

Así, pues, para todo $(x, y) \neq (0, -1)$ se tiene

$$g(x, y) = \frac{1}{4}(f(x, y) - 1)^2 - 1 \quad (4.80)$$

relación que, de hecho, es válida también en el punto $(0, -1)$, con lo que, en este ejemplo, vemos que f y g son funcionalmente dependientes (globalmente) en \mathbb{R}^2 .

Sean, en general, $f_1(x), \dots, f_m(x)$, $x = (x_1, \dots, x_n)$, m funciones de clase C^1 en un abierto $\Omega \subseteq \mathbb{R}^n$. Se dice que una función $f_k(x)$ *depende funcionalmente* en Ω de las funciones $f_i(x)$, $i = 1, \dots, m$, $i \neq k$, si existe una función φ de clase C^1 tal que $f_k(x) = \varphi(f_1(x), \dots, f_{k-1}(x), f_{k+1}(x), \dots, f_m(x))$ para todo $x \in \Omega$. Se dice que las funciones $f_1(x), \dots, f_m(x)$ son *funcionalmente dependientes* en Ω si una de ellas, al menos, depende funcionalmente de las restantes. En caso contrario, son *funcionalmente independientes*.

La condición necesaria de dependencia funcional se generaliza sin dificultad al caso general:

Teorema 4.16 Si $f_1(x), \dots, f_m(x)$ son funcionalmente dependientes en Ω , el rango de la matriz jacobiana

$$\left(\frac{\partial f_i}{\partial x_j}(x) \right), \quad i = 1, \dots, m, \quad j = 1, \dots, n$$

es estrictamente inferior a m en todo punto $x \in \Omega$.

Demostración. Si $m > n$, o sea, si el número de funciones es mayor que el de variables, el resultado es obvio, ya que el rango de la matriz jacobiana es $\leq n$. Supongamos, entonces, que $m \leq n$ y que una, al menos, de las funciones f_i depende de las restantes. Reordenando, si es necesario, las funciones, siempre podemos suponer que es $f_m(x)$ la que depende de las demás: $f_m(x) = \varphi(f_1(x), \dots, f_{m-1}(x))$. Derivando en esta igualdad se obtiene en todo $x \in \Omega$

$$\frac{\partial f_m}{\partial x_j} = \sum_{i=1}^{m-1} D_i \varphi \frac{\partial f_i}{\partial x_j}, \quad j = 1, \dots, n$$

Pero esto significa en la matriz jacobiana

$$\begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \vdots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{pmatrix}$$

que la última fila es combinación lineal de las demás, con lo que todos los menores de orden m (que han de contener las m filas) son nulos.

NOTA. Si $m = n$, el teorema nos dice que para que n funciones $f_1(x), \dots, f_n(x)$ de n variables sean funcionalmente dependientes en Ω , es necesario que el determinante jacobiano $Jf(x)$ (único menor de orden $m = n$) se anule en todo punto $x \in \Omega$.

Del teorema anterior se deduce inmediatamente el siguiente

Corolario 4.4 Si el rango de la matriz jacobiana

$$\left(\frac{\partial f_i}{\partial x_j}(x_0) \right), \quad i = 1, \dots, m, \quad j = 1, \dots, n$$

es igual a m en algún punto $x_0 \in \Omega$, entonces las funciones $f_1(x), \dots, f_m(x)$ son funcionalmente independientes en Ω . En particular, si el jacobiano de n funciones $f_1(x), \dots, f_n(x)$ respecto a $x = (x_1, \dots, x_n)$ es distinto de cero en algún punto $x_0 \in \Omega$, entonces las funciones $f_1(x), \dots, f_n(x)$ son funcionalmente independientes en Ω .

En el capítulo 9 demostraremos, generalizando trivialmente los argumentos expuestos anteriormente, la siguiente condición suficiente de dependencia local:

Teorema 4.17 Sean $f_i(x)$, $i = 1, \dots, m$, m funciones de clase C^1 en un entorno $U(x_0)$ del punto $x_0 \in \mathbb{R}^n$ y supongamos que

$$\text{rango} \left(\frac{\partial f_i}{\partial x_j}(x_0) \right) = r, \quad 0 < r < m$$

y que todos los menores de orden $r+1$ de la matriz jacobiana son idénticamente nulos en $U(x_0)$. Entonces, existe $\delta_0 > 0$ tal que las funciones $f_1(x), \dots, f_m(x)$ son funcionalmente dependientes en $B_{\delta_0}(x_0) \subseteq U(x_0)$.

NOTAS. 1. La hipótesis significa que hay un menor $J_r(x)$ de orden r de la matriz jacobiana que es distinto de cero en x_0 y también —por la continuidad de las derivadas $\partial f_i / \partial x_j$ — en todo x de un cierto entorno $B_r(x_0)$ de x_0 mientras que todos los menores de orden $r+1$ son idénticamente nulos en $B_r(x_0)$, es decir, que $\left(\frac{\partial f_i}{\partial x_j}(x) \right)$ tiene rango constante r siempre que $x \in B_r(x_0)$. Las r funciones que constituyen $J_r(x)$ son independientes en $B_{\delta_0}(x_0)$ y las $m-r$ restantes dependen funcionalmente de ellas.

2. Si $m > n$, o sea, si el número de funciones es mayor que el de variables, se verifica con seguridad

$$\text{rango} \left(\frac{\partial f_i}{\partial x_j}(x_0) \right) < m$$

ya que que el rango de la matriz jacobiana es $\leq n < m$. Así, pues, en el caso de más funciones que variables, si el rango de $\left(\frac{\partial f_i}{\partial x_j}(x_0) \right)$ se mantiene constante en un cierto entorno de x_0 , hay dependencia entre las funciones en un entorno $B_{\delta_0}(x_0)$ de x_0 .

De este resultado principal se derivan otras formulaciones que pueden ser útiles en la práctica:

Teorema 4.18 Sean $f_i(x)$, $i = 1, \dots, m$, m funciones de clase C^1 en un abierto conexo Ω de \mathbb{R}^n . Si el rango de la matriz jacobiana

$$\left(\frac{\partial f_i}{\partial x_j}(x) \right), \quad i = 1, \dots, m, \quad j = 1, \dots, n$$

es estrictamente inferior a m en todo punto $x \in \Omega$, entonces las funciones $f_1(x), \dots, f_m(x)$ son funcionalmente dependientes en un cierto subconjunto abierto de Ω .

En efecto, si todos los elementos de $\left(\frac{\partial f_i}{\partial x_j}(x) \right)$, o sea, todas las derivadas $\frac{\partial f_i}{\partial x_j}(x)$, se anulan para todo $x \in \Omega$, entonces $f_1(x), \dots, f_m(x)$ son constantes en Ω (que se supone conexo) y hay infinitas maneras de expresar cualquiera de las funciones en función de las demás.

Sea entonces $J_r(x)$ un menor de $Df(x)$ de orden máximo r , con $1 \leq r < m$, que no es idénticamente nulo en Ω . Se tiene, pues, $J_r(x_0) \neq 0$ para un cierto

$x_0 \in \Omega$ —y también, por la continuidad de las derivadas $\frac{\partial f_i}{\partial x_j}$, $J_r(x) \neq 0$ en $x \in B_\rho(x_0)$ para un cierto $\rho > 0$ — mientras que todos los menores de orden mayor que r son idénticamente nulos en Ω . Se aplica entonces el teorema anterior con $\Omega = U(x)$.

NOTA. Si $m = n$, este resultado dice que si el determinante jacobiano de n funciones $f_1(x), \dots, f_n(x)$ respecto a $x = (x_1, \dots, x_n)$ es igual a cero para todo $x \in \Omega$, entonces las funciones $f_1(x), \dots, f_n(x)$ son funcionalmente dependientes en un cierto subconjunto abierto de Ω . Si $m > n$, o sea, si el número de funciones es mayor que el de variables, se satisface obviamente la hipótesis ya que el rango de la matriz jacobiana es $\leq n < m$. Así, pues, en el caso de más funciones que variables, hay siempre dependencia entre las funciones, al menos en un subconjunto de Ω (supuesto éste contexto).

Teorema 4.19 Sean $f_i(x)$, $i = 1, \dots, m$, m funciones de clase C^1 en un abierto Ω de \mathbb{R}^n . Si

$$\text{rango} \left(\frac{\partial f_i}{\partial x_j}(x) \right) = r, \quad 0 < r < m$$

para todo $x \in \Omega$, entonces $f_1(x), \dots, f_m(x)$ son funcionalmente dependientes en un cierto entorno abierto de cada punto $x \in \Omega$.

En efecto, basta aplicar de nuevo el resultado principal con $U(x) = \Omega$.

NOTAS. 1. Fijado x , las r funciones que constituyen el menor de orden r distinto de cero en x son independientes en el correspondiente entorno de x y las $m - r$ restantes dependen funcionalmente de ellas en el entorno.

2. La aplicación más frecuente de este resultado, como de los anteriores, será al caso de varias funciones del mismo número de variables, o sea, cuando $m = n$.

3. Si $r = 0$, la hipótesis sería $\frac{\partial f_i}{\partial x_j}(x) = 0$ para todo $x \in \Omega$, lo que implica, si Ω es conexo, que $f_1(x), \dots, f_m(x)$ son constantes en Ω y que, en consecuencia, hay infinitas maneras de expresar cualquiera de las funciones en función de las demás.

EJEMPLO 5. Estudiemos la posible dependencia de las funciones

$$u(x, y, z) = x + y + z, \quad v(x, y, z) = x^2 + y^2 + z^2, \quad w(x, y, z) = xy + yz + zx$$

Se tiene

$$\frac{\partial(u, v, w)}{\partial(x, y, z)} = \begin{vmatrix} 1 & 1 & 1 \\ 2x & 2y & 2z \\ y+z & x+z & y+x \end{vmatrix} = 2(x+y+z) \begin{vmatrix} 1 & 1 & 1 \\ x & y & z \\ 1 & 1 & 1 \end{vmatrix} = 0$$

para todo $(x, y, z) \in \mathbb{R}^3$. El rango es 2 (considerando, por ejemplo, las dos primeras filas) en todo punto que no esté en la recta (x, x, x) , con lo que,

fuera de ellos, las dos primeras funciones, por ejemplo, son independientes y la tercera depende funcionalmente de ellas. Elevando al cuadrado la primera resulta

$$u^2 = x^2 + y^2 + z^2 + 2(xy + yz + zx) = v + 2w$$

relación válida, de hecho, para todo $(x, y, z) \in \mathbb{R}^3$.

Es interesante interpretar geoméricamente la condición de dependencia funcional. Retomemos para ello nuestro primer ejemplo: $u = x + y$, $v = x^2 + 2xy + y^2 = (x + y)^2 = u^2$. Consideremos un punto (x_0, y_0) arbitrario. El conjunto de nivel de u que pasa por dicho punto es la recta $x + y = x_0 + y_0$, mientras que el correspondiente conjunto de nivel de v que pasa por el mismo punto está formado por los (x, y) para los que se cumple: $(x + y)^2 = (x_0 + y_0)^2$, es decir: $x + y = \pm(x_0 + y_0)$, y está, pues, formado por la unión de las dos rectas paralelas $x + y = x_0 + y_0$ y $x + y = -(x_0 + y_0)$ (salvo que $x_0 + y_0 = 0$, en cuyo caso ambas coinciden). En cualquier caso, en un entorno de (x_0, y_0) los dos conjuntos de nivel coinciden.

En general, si $u = f(x, y)$ y $v = g(x, y)$ están ligadas funcionalmente por una relación del tipo $v = \varphi(u)$ en la que $\varphi' \neq 0$, y consideramos un (x_0, y_0) en el cual $f'_x \neq 0$ (por ejemplo), la teoría de funciones implícitas nos dice que el conjunto de nivel $f(x, y) = f(x_0, y_0)$ en un entorno de (x_0, y_0) es una curva $y = \phi(x)$ con $y_0 = \phi(x_0)$. Por otra parte, como $g(x, y) = \varphi(f(x, y))$ para todo (x, y) , vemos que $g(x, \phi(x)) = \varphi(f(x, \phi(x))) = \varphi(f(x_0, y_0)) = g(x_0, y_0)$. Es decir, la misma curva $y = \phi(x)$ está contenida en el conjunto de nivel $g(x, y) = g(x_0, y_0)$ de g que pasa por (x_0, y_0) . La monotonía estricta de φ nos indica, además, que, de hecho, coincide con este conjunto (al menos cerca de (x_0, y_0)), pues la ecuación $\varphi(x_0) = g(x_0, y_0)$ sólo se satisface para $x_0 = f(x_0, y_0)$. Lo mismo se podría decir si $f'_y(x_0, y_0) \neq 0$.

Vemos, pues, que si $\nabla f(x_0, y_0) \neq 0$ y $\varphi'(f(x_0, y_0)) \neq 0$, existe un entorno de (x_0, y_0) en el cual las curvas de nivel (que es lo que, con toda garantía, son los conjuntos de nivel en este caso) de f y de g que pasan por dicho punto coinciden.

Como $\nabla f(x_0, y_0)$ es ortogonal a la tangente a la curva de nivel de f en (x_0, y_0) y $\nabla g(x_0, y_0)$ tiene la misma propiedad respecto de la curva de nivel de g , y éstas coinciden, los dos vectores son ortogonales a una misma recta. Tratándose de vectores planos, ello implica que han de ser proporcionales, y entonces la matriz 2×2 cuyas filas son ∇f , ∇g (es decir, la matriz jacobiana de (f, g)) tiene determinante nulo en (x_0, y_0) .

Este argumento puede efectuarse para todo punto (x_0, y_0) en el cual $\nabla f \neq 0$ y $\varphi'(f(x_0, y_0)) \neq 0$. Si $\nabla f(x_0, y_0) = 0$, la anulación del determinante es obvia, y si $\varphi'(f(x_0, y_0)) = 0$, como $\nabla g(x, y) = \varphi'(f(x, y))\nabla f(x, y)$, tendríamos $\nabla g(x_0, y_0) = 0$ y, como antes, el determinante se anula.

Así pues, la condición puramente analítica $g(x, y) = \varphi(f(x, y))$ admite una interpretación geométrica (igualdad de los conjuntos de nivel) que, a su vez, implica la proporcionalidad de los gradientes y la consiguiente anulación de su determinante.

4.3 Notas al capítulo

1. Derivada general según un vector no necesariamente unitario.

Como ya se indicó en el texto, para definir la derivada según un vector no siempre se exige que éste sea unitario:

$$Df(x_0, \vec{u}) \text{ ó } D_{\vec{u}}f(x_0) \stackrel{\text{def}}{=} \lim_{t \rightarrow 0} \frac{f(x_0 + t\vec{u}) - f(x_0)}{t}$$

para cualquier $\vec{u} \in \mathbb{R}^n$ para el que el límite exista. En sí, esta extensión no presenta ninguna contradicción con la definición que nosotros utilizamos, a condición de que no se pretenda interpretar estas derivadas como *direccionales*: no porque no se trate de derivadas —que lo son—, sino porque el parámetro t deja de representar el “tamaño” (la longitud) real del incremento en las variables independientes, y pasa a representar un tamaño “relativo” al del vector \vec{u} . Esto se manifiesta en la relación $Df(x_0, \alpha\vec{u}) = \alpha Df(x_0, \vec{u})$, fácilmente demostrable, y que refleja el hecho de que el incremento $t(\alpha\vec{u})$ relativo al vector $\alpha\vec{u}$ tiene una longitud igual al correspondiente incremento $t\vec{u}$, multiplicado por α . En resumen, es aconsejable reservar la denominación “derivada direccional” a la derivada según un vector *unitario*, pues, de lo contrario, al haber infinitos vectores que marcan la misma dirección, habría infinitas derivadas en una dirección dada.

En el caso de que la función sea diferenciable en x_0 , esta derivada general viene dada por la misma fórmula de las derivadas direccionales:

$$Df(x_0, \vec{u}) = df(x_0; \vec{u}) = \nabla f(x_0) \cdot \vec{u}$$

con lo cual la derivada según un vector coincide exactamente con la diferencial, manteniéndose, como mucho, el matiz de hablar de “diferencial” cuando se piensa en la *aplicación lineal* $\vec{u} \rightarrow df(x_0, \vec{u})$, más que en su valor para un vector concreto \vec{u} .

2. Derivadas según una recta y según una curva.

Dada una recta l que pasa por (x_0, y_0) , se define a veces la derivada de f en (x_0, y_0) según la recta l , o en la dirección de la recta l , a su derivada en la dirección de un vector *unitario director* de la recta. Hay en esta definición una ambigüedad intrínseca, pues tanto \vec{u} como $-\vec{u}$ son vectores

directores de la misma recta, por lo que se debe suponer que el contexto aclara el sentido en que se supone recorrida la recta. En cualquier caso, dado que $D_{-\vec{u}}f(x_0, y_0) = -D_{\vec{u}}f(x_0, y_0)$, basta calcular la derivada según una de las direcciones para obtener inmediatamente la otra.

Por ejemplo, la recta $x + 2y - 3 = 0$ pasa por los puntos $(0, 3/2)$ y $(3, 0)$, de forma que $(3, -3/2)$ es un vector director suyo, cuyo módulo es $\frac{3}{2}\sqrt{5}$, y tanto $\vec{u} \stackrel{\text{def}}{=} (\frac{3}{2}\sqrt{5})^{-1}(3, -3/2) = (\frac{2}{\sqrt{5}}, -\frac{1}{\sqrt{5}})$ como $-\vec{u} = (-\frac{2}{\sqrt{5}}, \frac{1}{\sqrt{5}})$ son vectores directores de l . Otra forma, más sencilla, de obtener \vec{u} (ó $-\vec{u}$) es observar que la pendiente de l es $m = -1/2$ (coeficiente de x en la ecuación explícita $y = -(1/2)x + 3/2$) y un vector de dirección es $(1, m)$, cuyo vector normalizado es

$$\vec{u}_m \stackrel{\text{def}}{=} \left(\frac{1}{\sqrt{1+m^2}}, \frac{m}{\sqrt{1+m^2}} \right) \quad (4.81)$$

vector que coincide con \vec{u} , como se comprueba fácilmente al sustituir $m = -1/2$.

Análogamente, dada una curva expresada en forma explícita $y = g(x)$ que pase por (x_0, y_0) (es decir, tal que $f(x_0) = y_0$), la derivada de f en (x_0, y_0) según la curva $y = g(x)$ se define como la derivada según la recta tangente a la curva $y = g(x)$ en (x_0, y_0) , supuesta existente. Como para calcular la derivada de $f(x, y)$ según una recta lo único que hace falta conocer es su pendiente, y ésta vale $m = g'(x_0)$ en el caso de la tangente, vemos que la derivada buscada es la correspondiente a la dirección dada por

$$\vec{u}_m = \left(\frac{1}{\sqrt{1+g'(x_0)^2}}, \frac{g'(x_0)}{\sqrt{1+g'(x_0)^2}} \right)$$

En el caso n -dimensional, tanto la recta como la curva vendrían usualmente dadas en forma paramétrica, pues las versiones de tipo "cartesiano" son mucho más complicadas. La extensión de las nociones anteriores son, pues, directas, ya que bastaría con normalizar el vector director de la recta o de la tangente a la curva en \vec{x} .

En el mismo espíritu de la nota anterior, se produce frecuentemente una confusión acerca de estos conceptos, que consiste en que algunos autores definen la derivada según una recta l dada en forma paramétrica $\{x = x_0 + a_1 t, y = y_0 + a_2 t\}$ como la derivada de la función de t obtenida sustituyendo x y y por estos valores en en la fórmula que define $f(x, y)$: es decir, como la derivada respecto de t , evaluada en $t = 0$, de la función $\phi(t) = f(x_0 + a_1 t, y_0 + a_2 t)$. En otras palabras, se trataría de la derivada general de f según el vector director (a_1, a_2) . Por tanto, esta definición sólo coincide con la nuestra si el vector director (a_1, a_2) es unitario.

4.4 Problemas

- En los siguientes casos, calcular du/dt (o bien $\partial u/\partial t$ y $\partial u/\partial s$): (i) como derivada de una función compuesta y (ii) por sustitución directa, comparando los resultados.
 - $u = x^2 + y^2$; $x = \cos t$, $y = \sin t$.
 - $u = e^x y - \ln(y + z)$; $x = t^2$, $y = 0$, $z = t$.
 - $u = x + y^2 - z$; $x = t^2 + s^2$, $y = t - s$, $z = t$.
 - $u = x + y^2$; $x = p - q$, $y = p + q^2$; $p = t$, $q = t + s^2$.
 - $u = x^2 + t$; $x = \cos t$ (Atención a los distintos significados de du/dt y $\partial u/\partial t$).
 - $u = x - y^2 + t$; $y = -t$ (¿De cuántas maneras se puede interpretar $\partial u/\partial t$?).

En los problemas (c) y (d), si partimos de $t = 10$, $s = 5$, ¿cuál es el efecto aproximado sobre u de incrementar t en una cantidad $\Delta t = 0,1$? ¿Y si el incremento de t fuese del 2%? En general, si Δt es positivo (y suficientemente pequeño) ¿puede asegurarse que Δu será también positivo?

- Calcular dy/dx , donde $y = u^v$, $u = u(x)$, $v = v(x)$.
- Calcular la derivada de $u = x^2 + y^2 x$ a lo largo de la curva $y = x^3$. Particularizar en el punto $(1, 1)$. Lo mismo a lo largo de la curva $x = t^2$, $y = \sin t$, particularizando en el punto $(\pi^2, 0)$.
- Calcular la velocidad con que varía el volumen de un cilindro de revolución sabiendo que, en el instante considerado, el radio de la base vale 10 cm y crece a la velocidad de 2 cm/seg, y que la altura h vale 15 cm y crece a una velocidad de 1 cm/seg.
- Dada una función derivable $\varphi: \mathbb{R} \rightarrow \mathbb{R}$ de una variable, se pide:
 - Calcular las derivadas parciales de las funciones:
 - $f(x, y) = x\varphi(y) - y\varphi(x)$
 - $f(x, y) = x \sin \varphi(y)$
 - $f(x, y) = \varphi(x - y)$

- Probar que la función $f(x, y) = \varphi\left(\frac{y}{x}\right)$ satisface la ecuación

$$xf_x + yf_y = 0.$$

6. Dada la función $f(x, y) = x + y^3x^2$ se pide:
- Obtener su derivada en el punto $(1, 2)$ en la dirección de ángulo: (i) 30° ; (ii) 60° ; (iii) 270° ; (iv) π (radianes).
 - Obtener su derivada en el punto $(1, 2)$ en la dirección dada por: (i) la recta $x + y = 3$; (ii) la curva $x^2 + y^2 = 5$ (véase la nota 2 de final del capítulo 4).
 - Obtener su derivada en el punto $(1, 2)$ en la dirección dada por el vector: (i) $(1/\sqrt{2}, 1/\sqrt{2})$; (ii) $(1, -2)$.
7. Calcular los puntos (x, y, z) en los cuales la derivada de la función $f(x, y, z) = xy - z^2$ en la dirección dada por el vector $(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{-1}{\sqrt{3}})$ es nula.
8. ¿Cuáles de las siguientes funciones son homogéneas?

$$\begin{array}{ll} a) f(x, y) = x^2 + xy & b) f(x, y) = \frac{x^2 + xy}{x + y} \\ c) f(x, y) = \operatorname{sen} \frac{y}{x} & d) f(x, y) = x + xy \end{array}$$

Compruébese la fórmula de Euler para aquellas que lo sean.

9. (a) Sea $f(x, y)$ homogénea de grado k . Compruébese, directamente y mediante la fórmula de Euler, que $[f(x, y)]^n$ es de grado kn .
- (b) Compruébese que

$$u(x, y, z) = \left(\frac{x^2 + y^2 - z^2}{x - z} \right)^n$$

satisface

$$xu_x + yu_y + zu_z = nu.$$

10. Se consideran las variables x, y ligadas por la relación

$$y - xy^2 = 0 \quad (*)$$

- Comprobar que en un entorno de $(1, 1)$ la ecuación $(*)$ define a y como función (implícita) de x . Sea $\varphi(x)$ dicha función. ¿Cuánto vale $\varphi(1)$?
- Calcular la derivada primera de $\varphi(x)$.
- Escribir la ecuación de la recta tangente en $(1, \varphi(1))$ a la curva representativa de $y = \varphi(x)$.

- (d) Escribiendo la ecuación (*) en la forma $y(xy - 1) = 0$, dibujar la curva por ella representada e interpretar los resultados antes obtenidos.
- (e) ¿Pueden efectuarse cálculos análogos en el punto $(1,0)$? ¿Con qué resultado? ¿Y en el $(0,1)$?
11. Estudiar la aplicabilidad del teorema de la función implícita a las ecuaciones siguientes:

$$\begin{array}{ll} a) \quad ax^2 + 2bxy + cy^2 + 1 = 0 & b) \quad y + \cos(x+y) = 0 \\ c) \quad \sin xy - x = 0 & d) \quad x^y = y^x \\ e) \quad (x^2 + y^2)^2 - 2a^2(x^2 - y^2) = 0 & \text{(lemniscata)} \end{array}$$

Calcular en cada caso dy/dx . Comparar con los resultados que se obtendrían para la función explícita $y = \varphi(x)$ cuando ésta sea posible de obtener. En el caso e), localizar y clasificar los puntos críticos de $y = \varphi(x)$.

12. Dada la función $F(x, y, z) = xy + xz + yz - e^z$, se pide:

- (a) ¿Es diferenciable en el punto $(1, 1, 0)$?
- (b) Calcular aproximadamente $F(0,99; 1,01; 0,05)$
- (c) Probar que la ecuación $F(x, y, z) = 0$ define implícitamente a z como función de (x, y) , $z = f(x, y)$, en un entorno del punto $(1, 1)$.
- (d) Calcular el valor de las derivadas parciales de $f(x, y)$ en el punto $(1, 1)$.
- (e) Escribir la ecuación del plano tangente a la gráfica de $z = f(x, y)$ en el punto $(1, 1, 0)$.
- (f) ¿Cuál es la pendiente de esa gráfica en la dirección que va de $(1, 1)$ a $(0, 0)$?

13. Dada la función $F(x, y, z) = x + y + z - e^{xyz}$, se pide

- (a) Demostrar que $F(x, y, z) = 0$ define una única función implícita $z = f(x, y)$ en el entorno del punto $(0, \frac{1}{2})$.
- (b) Escribir la ecuación del plano tangente a la gráfica de $z = f(x, y)$ en el punto $(0, \frac{1}{2}, \frac{1}{2})$.
- (c) ¿Cuál es la pendiente de esa gráfica en la dirección que va del punto $(0, \frac{1}{2})$ al punto $(\frac{1}{2}, 1)$?

- (d) Considerando el mismo punto $(0, \frac{1}{2}, \frac{1}{2}) \in \mathbb{R}^3$, determinar si la ecuación $F(x, y, z) = 0$ define también a y como función de (x, z) en el entorno de $(0, \frac{1}{2})$ y a x como función implícita de (y, z) en el entorno de $(\frac{1}{2}, \frac{1}{2})$. Si la respuesta es positiva, calcular las derivadas primeras de dichas funciones implícitas en los puntos indicados.

14. Dada la ecuación

$$x - y^2 + 2z - 2xyz + z^3 = 0 \quad (*)$$

¿Qué variables pueden despejarse en función de las demás en un entorno de $(0, 0, 0)$? Supuesto que pueda expresarse z como función de (x, y) , calcular z_x, z_y y particularizarlas para el punto $x = 0, y = 0$. ¿Cuál es el valor de z correspondiente a dichos valores de x, y ? ¿Cuál es la aproximación lineal de z como función de (x, y) en el punto $(0, 0)$? Compararla con la que se obtiene aproximando $(*)$ por eliminación de todos los términos de grado mayor que 1, es decir despejando z de la ecuación $x + 2z = 0$.

15. Supuesto que la ecuación

$$F(x, y, z) = 0$$

permite definir cada una de las variables en función de las otras dos, comprobar que se verifica

$$\left(\frac{\partial x}{\partial y}\right)_z \left(\frac{\partial y}{\partial z}\right)_x \left(\frac{\partial z}{\partial x}\right)_y = -1$$

Establecer otra relación del mismo tipo.

16. Dada la función

$$f(x, y) = e^{1-(x/y)} - \cos\left(\frac{y}{x}\right)$$

se pide:

- Estudiar su dominio de definición.
- Analizar si dicha función es homogénea. En caso de serlo, ¿cuál es su grado?
- Determinar, sin calcular el gradiente de la función, el valor de la expresión

$$xf_x(x, y) + yf_y(x, y).$$

- (d) ¿Define la ecuación $f(x, y) = k$, para algún k , a la variable y como función de la variable x en un entorno de $x = 1, y = 1$? En caso afirmativo, obtener su derivada dy/dx y la aproximación lineal de y como función de x en el punto $x = 1$.

17. (a) Se considera la función

$$u = f(x, y) = x^3 - \frac{y}{x}$$

Supongamos que x e y dependen de la variable p mediante las fórmulas

$$x = e^p, \quad y = e^{-p}.$$

Averiguar, sin sustituir directamente p , si u es función creciente de p .

- (b) Se considera la función

$$u = f(x, y) = g(x) - \frac{y}{x}$$

donde $g(x)$ es una función de una variable tal que $g'(x) > 0$ para todo x . ¿Puede asegurarse que u es creciente como función de p , supuesto que x, y están relacionadas con p por las fórmulas anteriores?

18. Dada la ecuación

$$e^{-z} - \frac{1-x}{y} - yz = 0 \quad (*)$$

se pide:

- Comprobar que $(*)$ define a z como función implícita de (x, y) en un entorno de $(0, 1, 0)$.
- Obtener la aproximación lineal de z como función de (x, y) en el punto $(0, 1)$. Si a partir de $x = 0, y = 1$ damos unos incrementos $\Delta x = 0, 1, \Delta y = -0, 3$, calcular aproximadamente el incremento de z correspondiente.
- Despejar efectivamente z en función de (x, y) , obtener su aproximación lineal en $(0, 1)$ y compararla con la anteriormente obtenida.

19. Si en la función $z = f(x, y)$ se hace un cambio a las variables (s, t) relacionadas con (x, y) mediante las fórmulas

$$x + y = s, \quad x + 2y = t$$

se obtiene la nueva función (compuesta) $z = F(s, t)$ (véase el ejemplo 3 del apartado 4.1). Mediante el cálculo de z_x, z_y en función de z_s, z_t , obtener las soluciones de la ecuación en derivadas parciales

$$2z_x - z_y = 0.$$

20. (Del libro *Cálculo Integral* de P. Puig Adam) La derivación bajo el signo integral permite a veces calcular integrales definidas más fácilmente que por otros métodos. Por ejemplo, derivando respecto a t en el resultado conocido

$$\int_0^b \cos tx \, dx = \frac{1}{t} \sin tb$$

se obtiene

$$-\int_0^b x \sin tx \, dx = -\frac{1}{t^2} \sin tb + \frac{b}{t} \cos tb \quad (*)$$

y, de aquí, el valor de

$$\int_0^b x \sin tx \, dx$$

para el valor de t que interese. Derivar de nuevo en $(*)$ para obtener el valor de

$$\int_0^b x^2 \cos tx \, dx$$

y así sucesivamente.

Las integrales del tipo

$$\int \frac{dx}{(x^2 + a^2)^k}$$

se resuelven por fórmulas de reducción (de k a $k-1, \dots$) un tanto complicadas. Si se trata de la integral definida

$$\int_0^t \frac{dx}{(x^2 + t^2)^k}$$

se puede calcular más fácilmente empezando por la integral para $k = 1$

$$\int_0^t \frac{dx}{(x^2 + t^2)} = \frac{1}{t} \left[\arctan \frac{x}{t} \right]_0^t = \frac{\pi}{4t}$$

y derivando sucesivamente respecto a t :

$$\frac{1}{2t^2} - \int_0^t \frac{2t \, dx}{(x^2 + t^2)^2} = -\frac{\pi}{4t^2}, \quad \int_0^t \frac{dx}{(x^2 + t^2)^2} = \frac{1}{8t^3}(\pi + 2)$$

Derivar de nuevo respecto a t en esta última expresión para obtener

$$\int_0^t \frac{dx}{(x^2 + t^2)^3}$$

21. (a) Calcular la integral

$$\int_0^1 \frac{\ln(1+x)}{1+x^2} \, dx$$

mediante la introducción de un parámetro t :

$$G(t) = \int_0^1 \frac{\ln(1+tx)}{1+x^2} \, dx$$

y derivando respecto a t .

- (b) Calcular por derivación respecto al parámetro t las integrales:

$$\begin{aligned} \text{(i)} \quad & \int_0^{\pi} \ln(1 + t \cos x) \, dx & \text{(ii)} \quad & \int_0^{\pi/2} \frac{\ln(1 + t \sin x)}{\sin x} \, dx \\ \text{(iii)} \quad & \int_0^{\pi/2} \frac{\ln(1 + t \sin^2 x)}{\sin^2 x} \, dx \end{aligned}$$

22. Calcular las derivadas en $t = 0$ de las funciones $x = x(t)$, $y = y(t)$ definidas implícitamente por el sistema de ecuaciones

$$\begin{cases} x - y^2 - t = 0 \\ y + xy - xt = 0 \end{cases}$$

23. Calcular la matriz jacobiana asociada a la transformación lineal

$$u = x + 2y, \quad v = x - y$$

Compararla con su matriz asociada en cuanto transformación lineal de \mathbb{R}^2 en \mathbb{R}^2 . Obtener la transformación inversa y su correspondiente matriz jacobiana, comprobando que se trata de la matriz inversa de la anterior. ¿Qué se puede decir para una matriz A $n \times n$?

24. Probar que el sistema de ecuaciones

$$\begin{aligned}x^2 + y^2 + z^2 &= 1 \\ 3x - 2y + 4z &= 4\end{aligned}$$

define a y, z como funciones implícitas de x en un entorno de $(0, 0, 1)$ y calcular $dy/dx, dz/dx$.

25. Dado el sistema de ecuaciones

$$\begin{cases} x - y^2 + 2 \cos(xy) - e^z = 0 \\ x^2 + 2xy - \sin z = 0 \end{cases}$$

discutir la posibilidad de que dicho sistema defina a x, z como funciones implícitas de y , y calcular $dx/dy, dz/dy$ en caso de que sea posible.

26. (a) Se considera la transformación a coordenadas polares

$$x = r \cos \theta, \quad y = r \sin \theta$$

Calcular la matriz jacobiana $D_{(r,\theta)}(x,y)$ de (x,y) respecto a (r,θ) y comprobar que

$$\left. \frac{\partial(x,y)}{\partial(r,\theta)} \right|_{(r_0,\theta_0)} = r_0$$

- (b) ¿Cuándo se puede formar la función inversa $(r(x,y), \theta(x,y))$? Calcular también $D_{(x,y)}(r,\theta)$ y comprobar que es la matriz inversa de la anterior.

27. Se considera la transformación a coordenadas esféricas (véase el capítulo 1)

$$x = \rho \sin \phi \cos \theta, \quad y = \rho \sin \phi \sin \theta, \quad z = \rho \cos \phi$$

Calcular la matriz jacobiana asociada a esta transformación y comprobar que

$$\frac{\partial(x,y,z)}{\partial(\rho,\phi,\theta)} = \rho^2 \sin \phi$$

Estudiar cuándo dicha transformación es invertible.

28. Dadas las funciones

$$u(x,y) = x + y, \quad v(x,y) = x^2 + 2xy + y^2$$

se pide:

- (a) Obtener sus familias de curvas de nivel.
- (b) Calcular $\nabla u, \nabla v$. Comprobar que ∇u está siempre alineado con ∇v . ¿Qué implica esto en cuanto al jacobiano $\partial(u, v)/\partial(x, y)$?
- (c) Estudiar la dependencia funcional de $u(x, y)$ y $v(x, y)$.

29. Analizar la dependencia funcional de las funciones

$$u(x, y, z) = x(y - z), \quad v(x, y, z) = y(z - x), \quad w(x, y, z) = z(x - y)$$

Calcular $u + v + w$ e interpretar el resultado.

30. Dadas las funciones

$$u(x, y) = x + y, \quad v(x, y) = x - y, \quad w(x, y) = y^2$$

probar que $w(x, y)$ se puede expresar como función de $u(x, y)$ y $v(x, y)$.

31. Analizar la dependencia funcional de las funciones

$$\begin{aligned} u(x, y, z) &= x^3 + y^3 + z^3 - 3xyz, & v(x, y, z) &= x + y + z, \\ w(x, y, z) &= x^2 + y^2 + z^2 - xy - zy - zx \end{aligned}$$

Encontrar la relación entre u, v, w .

32. Una *ecuación diferencial ordinaria* es una relación que ha de satisfacer una "función incógnita" de una variable $x(t)$ junto con su derivada $x'(t)$:

$$x' = f(t, x)$$

(Esta es la *forma normal* de una ecuación diferencial, con la derivada expresada como función explícita de la variable independiente t y la función x ; la definición de solución de una ecuación diferencial ordinaria es, como en el problema 15 del capítulo anterior, obvia). La ecuación se dice que es de *variables separables* si la función del segundo miembro es el producto de una función sólo de t por una función sólo de x . Son, por tanto, de la forma

$$\frac{dx}{dt} = g(t)h(x) \quad (*)$$

donde g y h se suponen continuas en los intervalos (a, b) y (c, d) respectivamente. Se trata de probar el siguiente teorema de existencia de soluciones para el *problema de valor inicial* asociado a la ecuación (*):

Teorema Sean $t_0 \in (a, b)$, $x_0 \in (c, d)$. Entonces se tiene:

- (a) Si $h(x_0) = 0$, la función constante $x(t) = x_0$ para todo $t \in (a, b)$ satisface la ecuación (*) y la condición inicial $x(t_0) = x_0$.
- (b) Si $h(x_0) \neq 0$, existe un intervalo $I \subseteq (a, b)$, con $t_0 \in (a, b)$, tal que la función $x(t)$ definida implícitamente por

$$\int_{x_0}^x \frac{dz}{h(z)} = \int_{t_0}^t g(s) ds$$

es la única solución de (*) definida en I que satisface la condición inicial $x(t_0) = x_0$.

(Indicación: Considérense las primitivas

$$H(x) = \int_{x_0}^x \frac{dz}{h(z)} \text{ y } G(t) = \int_{t_0}^t g(s) ds$$

y la ecuación

$$H(x) - G(t) = 0$$

Este teorema justifica el nombre de *método de separación de variables* que se da al proceso por el que se resuelven en la práctica estas ecuaciones: para encontrar las soluciones de

$$\frac{dx}{dt} = g(t)h(x)$$

se "separan las variables"

$$\frac{dx}{h(x)} = g(t)dt$$

y, tomando integrales indefinidas en ambos miembros,

$$\int \frac{dx}{h(x)} = \int g(t)dt + C$$

con C constante arbitraria, se obtiene una ecuación

$$H(x) = G(t) + C$$

que define implícitamente a las soluciones de la ecuación (*), excepto a las posibles soluciones constantes $x(t) \equiv x_0$ con $h(x_0) = 0$ (que se detectan directamente a partir de la propia ecuación). Aquí, H y G son, respectivamente, determinadas primitivas de $1/h(x)$ y $g(t)$. Para cada C tal que la ecuación $H(x) = G(t) + C$ tenga al menos una solución

(t, x) , se tendrá una o más soluciones de la ecuación $(*)$ (una sola si ha de satisfacer además la condición inicial $x(t_0) = x_0$ con $h(x_0) \neq 0$, y nos restringimos a un intervalo suficientemente pequeño centrado en $t_0 \in (a, b)$).

Sea, por ejemplo, la ecuación diferencial

$$\frac{dx}{dt} = 2(1+x^2)t$$

Procediendo como se ha indicado, las soluciones estarán dadas por

$$\int \frac{dx}{1+x^2} = \int 2t \, dt + C$$

es decir

$$\arctan x = t^2 + C$$

o sea

$$x(t) = \tan(t^2 + C)$$

con C constante arbitraria (hay toda una familia de soluciones —la *solución general*— dependiente de una constante arbitraria). Si se pregunta por la solución concreta que satisface la condición inicial prefijada $x(0) = 0$, se impone esta condición en la fórmula de la solución general:

$$0 = x(0) = \tan(0 + C) = \tan C$$

y queda determinado el valor de C que le corresponde: $C = 0$, con lo que la solución pedida es

$$x(t) = \tan t$$

Resolver los siguientes problemas de valor inicial:

$$a) \begin{cases} x' = tx \\ x(0) = 1 \end{cases} \quad b) \begin{cases} x' = \frac{t}{x} \\ x(1) = 2 \end{cases} \quad c) \begin{cases} x' = -\frac{t^2 x}{1+t^3} \\ x(0) = 1 \end{cases}$$

Resolver el problema de valor inicial genérico para la ecuación logística:

$$\begin{cases} x' = ax - bx^2 \\ x(0) = x_0 \end{cases}$$

33. Puede ocurrir que una ecuación diferencial que en principio no sabemos resolver se transforme, mediante un cambio de variable, en una ecuación de variables separables: dada la ecuación $x' = f(t, x)$, el cambio de variables dado por

$$u(t) = H(t, x(t))$$

(que se expresa abreviadamente por $u = H(t, x)$) nos lleva, por la regla de la cadena, a la siguiente ecuación diferencial en la nueva variable dependiente u :

$$\begin{aligned} u'(t) &= H_t(t, x(t)) + H_x(t, x(t)) \cdot x'(t) = \\ &= H_t(t, x(t)) + H_x(t, x(t)) \cdot f(t, x(t)) = \\ &= H_t(t, K(t, u(t))) + H_x(t, K(t, u(t))) \cdot f(t, K(t, u(t))) \end{aligned}$$

(abreviadamente $u' = H_t(t, K(t, u)) + H_x(t, K(t, u))f(t, K(t, u))$, donde $x = K(t, u)$ es el cambio de variables inverso, es decir, el resultado de despejar x en función de u en la ecuación $u = H(t, x)$).

- (a) Probar que una ecuación diferencial de la forma

$$x' = f(at + bx + c)$$

se transforma en una ecuación de variables separables mediante el cambio $u = at + bx$.

- (b) Una ecuación diferencial $x' = f(t, x)$ se dice que es *homogénea* si la función $f(t, x)$ es homogénea de grado cero. Esto implica que $f(t, x) = F\left(\frac{x}{t}\right)$ para una cierta función F de una variable (¿por qué?). Probar que una ecuación diferencial homogénea se reduce a una de variables separables mediante el cambio

$$u = \frac{x}{t}$$

- (c) Resolver las siguientes ecuaciones

$$(i) x' = \frac{x}{t} + \frac{t}{x} \quad (ii) x' = \frac{x+t}{x-t} \quad (iii) x' = \frac{x + \sqrt{t^2 + x^2}}{t}$$

Determinar la solución de (iii) que satisface la condición inicial $x(1) = 0$.

34. Dada la ecuación diferencial

$$\frac{dx}{dt} = \frac{at + bx + c}{pt + qx + r}$$

probar que se transforma en la ecuación homogénea

$$\frac{dX}{dT} = \frac{aT + bX}{pT + qX}$$

mediante el cambio $X = x - h$, $T = t - k$, siendo (h, k) el punto de corte de las rectas $at + bx + c = 0$, $pt + qx + r = 0$. Tratar también el caso especial en que $aq - bp = 0$, o sea, cuando dichas rectas son paralelas, mediante el cambio $u = at + bx$.

Aplicar esto a la resolución de las ecuaciones:

$$(i) \ x' = \frac{2t - 5x + 3}{2t - 4x - 6} \quad (ii) \ x' = \frac{x + t - 2}{x - t} \quad (iii) \ x' = -\frac{8t + 4x + 1}{4t + 2x + 1}$$

Capítulo 5

Derivadas sucesivas. Fórmula de Taylor

Continuamos en este capítulo con la extensión a varias variables de conceptos y resultados bien conocidos del cálculo diferencial de funciones de una variable. Las derivadas segundas de una función de varias variables son, simplemente, las derivadas de las derivadas primeras, y así sucesivamente. Eso sí, la multiplicidad de derivadas de un determinado orden plantea la cuestión, nueva respecto al cálculo de funciones de una variable, de estudiar qué relaciones existen entre ellas. La fórmula de Taylor es un instrumento fundamental para el estudio cualitativo y cuantitativo de una función y su extensión a funciones de varias variables se basa directamente en el correspondiente resultado para funciones de una variable. En una primera aproximación, puede limitarse la lectura de este capítulo a los apartados 1, 2 y 3.

5.1 Derivadas sucesivas. Funciones de clase C^p .

Recordemos que si $y = f(x)$ es una función de una variable definida en un intervalo I , y suponemos que es derivable en todos los puntos de I , llamábamos *función derivada* a la función $x \mapsto f'(x)$. Si esta función admite, a su vez, derivada en un punto $x_0 \in I$, la llamamos *derivada segunda* de f en x_0 y la denotamos por $f''(x_0)$ ó $D^2f(x_0)$:

$$f''(x_0) \stackrel{\text{def}}{=} (f')'(x_0) \text{ ó } D^2f(x_0) = (D(Df))(x_0)$$

Si f'' está bien definida en todo I , llamamos *función derivada segunda* a $x \mapsto f''(x)$, función cuya derivada, si existe, denominamos *derivada tercera*, y así sucesivamente. En general, la *derivada n -ésima* se denota $f^{(n)}(x)$, $D^n f(x)$ ó $d^n f/dx^n$.

La extensión a funciones de dos o más variables del concepto de derivadas sucesivas es inmediata.

Derivadas parciales segundas

Sea $f(x, y)$ una función definida en un conjunto abierto $D \subset \mathbb{R}^2$ tal que f'_x y f'_y existen en todo punto de D . Están, entonces, definidas las funciones derivadas parciales (primeras)

$$D \ni (x, y) \mapsto f'_x(x, y)$$

$$D \ni (x, y) \mapsto f'_y(x, y)$$

Las derivadas parciales de estas funciones en un punto (x_0, y_0) , si existen, son las *derivadas parciales segundas* de f , a saber:

- $(f'_x)'_x(x_0, y_0)$: derivada (parcial) segunda de f respecto a x (dos veces); se denota de cualquiera de las formas siguientes:

$$f''_{xx}(x_0, y_0), f''_{xx}(x_0, y_0), f_{xx}(x_0, y_0), D^2_{xx}f(x_0, y_0), D^2_{xx}f(x_0, y_0), D^2_{11}f(x_0, y_0), f_{11}(x_0, y_0), \frac{\partial^2 f}{\partial x^2}(x_0, y_0)$$

- $(f'_y)'_y(x_0, y_0)$: derivada segunda respecto a y (dos veces); se denota por

$$f''_{yy}, f''_{yy}, \dots, D^2_{yy}f, f_{yy}, \frac{\partial^2 f}{\partial y^2}$$

(evaluando en (x_0, y_0)).

- $(f'_y)'_x(x_0, y_0)$: derivada segunda respecto a x y respecto a y :

$$f''_{xy}, f_{xy}, D^2_{xy}f, D^2_{12}f, f_{12}, \frac{\partial^2 f}{\partial x \partial y}$$

(evaluando en (x_0, y_0)).

- $(f'_x)'_y(x_0, y_0)$: derivada segunda respecto a y y respecto a x :

$$f''_{yx}, f_{yx}, D^2_{yx}f, D^2_{21}f, f_{21}, \frac{\partial^2 f}{\partial y \partial x}$$

(evaluando en (x_0, y_0)).

Si estas derivadas están definidas en todo punto $(x, y) \in D$, entonces $(x, y) \mapsto \frac{\partial^2 f}{\partial x^2}(x, y)$, etc., son las correspondientes **funciones derivadas segundas** de f . Las dos últimas son las derivadas *mixtas* o *cruzadas* de f , y plantean la cuestión, sin equivalente en el cálculo de funciones de una variable, de estudiar qué relación existe entre ellas: ¿es posible cambiar el orden de derivación?

EJEMPLOS.

1. Sea $f(x, y) = x^5 y^4 + y e^{2x}$. Entonces:

$$\frac{\partial}{\partial x}(x^5 y^4 + y e^{2x}) = 5x^4 y^4 + 2y e^{2x}$$

$$\frac{\partial}{\partial y}(x^5 y^4 + y e^{2x}) = 4x^5 y^3 + e^{2x}$$

$$\frac{\partial^2}{\partial x^2}(x^5 y^4 + y e^{2x}) = \frac{\partial}{\partial x}(5x^4 y^4 + 2y e^{2x}) = 20x^3 y^4 + 4y e^{2x}$$

$$\frac{\partial^2}{\partial y^2}(x^5 y^4 + y e^{2x}) = \frac{\partial}{\partial y}(4x^5 y^3 + e^{2x}) = 12x^5 y^2$$

$$\frac{\partial^2}{\partial x \partial y}(x^5 y^4 + y e^{2x}) = \frac{\partial}{\partial y}(5x^4 y^4 + 2y e^{2x}) = 20x^4 y^3 + 2e^{2x}$$

$$\frac{\partial^2}{\partial y \partial x}(x^5 y^4 + y e^{2x}) = \frac{\partial}{\partial x}(4x^5 y^3 + e^{2x}) = 20x^4 y^3 + 2e^{2x}$$

2. Sea $f(x, y) = e^x y^2$. Entonces:

$$f'_x = D_x(e^x y^2) = e^x y^2, \quad f'_y = D_y(e^x y^2) = 2e^x y$$

$$f''_{xx} = D_x D_x f = D_x(e^x y^2) = e^x y^2, \quad f''_{xy} = D_y D_x f = D_y(e^x y^2) = 2e^x y$$

$$f''_{yx} = D_x D_y f = D_x(2e^x y) = 2e^x y, \quad f''_{yy} = D_y D_y f = D_y(2e^x y) = 2e^x$$

Observamos en estos dos ejemplos que $f''_{xy} = f''_{yx}$, es decir, que las dos derivadas mixtas coinciden y que, por tanto, *no importa el orden de derivación*. Esto se cumple bajo hipótesis muy generales que se dan en la gran mayoría de las aplicaciones. En efecto, se tiene el siguiente teorema que demostraremos al final del apartado:

Teorema 5.1 Si f''_{xy} y f''_{yx} existen en un entorno de (x_0, y_0) y son continuas en (x_0, y_0) , entonces coinciden en dicho punto:

$$f''_{xy}(x_0, y_0) = f''_{yx}(x_0, y_0)$$

La extensión a funciones $f(x_1, \dots, x_n)$ de n variables es inmediata: Sea $D \subset \mathbb{R}^n$ un conjunto abierto y sea $f: D \rightarrow \mathbb{R}$ una función cuyas derivadas parciales están definidas en todo D . La derivada (parcial) segunda respecto a x_i y x_j en un punto $x_0 \in D$ es la derivada parcial respecto a x_j de la función derivada parcial respecto a x_i de f , y se denota por cualquiera de los siguientes símbolos

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(x_0), D^2_{x_i x_j} f(x_0), D^2_{ij} f(x_0), f''_{x_i x_j}(x_0), f''_{ij}(x_0), f_{x_i x_j}(x_0), f_{ij}(x_0)$$

Es decir:

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(x_0) \stackrel{\text{def}}{=} \frac{\partial}{\partial x_j} \left(\frac{\partial f}{\partial x_i} \right)(x_0), \quad D^2_{x_i x_j} f(x_0) \stackrel{\text{def}}{=} D_{x_j}(D_{x_i} f)(x_0),$$

$$f''_{x_i x_j}(x_0) \stackrel{\text{def}}{=} (f'_{x_i})'_{x_j}(x_0), \text{ etc.}$$

Si esta derivada está definida para todo $x \in D$, entonces $x \mapsto f''_{x_i x_j}(x)$ se denomina **función derivada segunda respecto de x_i y x_j** .

Puesto que en la determinación de $\frac{\partial^2 f}{\partial x_i \partial x_j}(x_0)$ y $\frac{\partial^2 f}{\partial x_j \partial x_i}(x_0)$ las variables distintas a x_i y x_j permanecen constantes, o sea, se calculan las derivadas segundas de la función de dos variables

$$(x_i, x_j) \mapsto f(x_1^0, \dots, x_i, \dots, x_j, \dots, x_n^0)$$

en el punto (x_i^0, x_j^0) , el teorema 5.1 se aplica directamente para obtener que,

si $\frac{\partial^2 f}{\partial x_i \partial x_j}$ y $\frac{\partial^2 f}{\partial x_j \partial x_i}$ están definidas en un entorno de x_0 y son continuas en x_0 , entonces

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(x_0) = \frac{\partial^2 f}{\partial x_j \partial x_i}(x_0)$$

En general, trabajaremos con funciones derivadas parciales definidas en todo un conjunto abierto D . Conviene, por ello, introducir la siguiente

Definición 5.1 Si en un abierto $D \subset \mathbb{R}^n$ existen y son continuas todas las derivadas parciales primeras y segundas de la función f , se dice que f es **dos veces continuamente diferenciable** y también que es de **clase dos** (o de **clase C^2**) en D . El conjunto de tales funciones se denota por $C^2(D)$.

(Nótese que de la continuidad de las derivadas segundas se deduce, por el teorema 3.2, la continuidad de las derivadas primeras y de la propia función f .)

De acuerdo con el teorema 5.1, si $f \in C^2(D)$, entonces

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(x) = \frac{\partial^2 f}{\partial x_j \partial x_i}(x), \quad x \in D \quad (5.1)$$

para todo $i, j = 1, \dots, n$.

La igualdad (5.1) de las funciones derivadas segundas cruzadas será, como decíamos, lo usual en las aplicaciones.

Aplicación: Funciones potenciales. Campos gradiente

Una función continua $f(x)$ de una variable definida en un intervalo I siempre posee una *función primitiva*, o sea, una función $F(x)$ tal que $F'(x) = f(x)$ (y, como sabemos, si $F(x)$ es función primitiva de $f(x)$, también lo es $F(x) + C$ para cualquier valor de la constante C). Basta considerar, en efecto,

$$F(x) = \int_{x_0}^x f(t) dt$$

siendo x_0 un cierto punto de I .

Si nos planteamos la misma cuestión para funciones de dos o más variables, no tiene sentido preguntarse por la primitiva de una sola función $f(x, y)$, pues una función de dos variables tiene *dos* derivadas primeras. Es claro que la cuestión análoga habrá de ser: dado un par de funciones $f(x, y)$, $g(x, y)$ en un conjunto abierto D ¿existe una función $u(x, y)$ tal que

$$u_x(x, y) = f(x, y), \quad u_y(x, y) = g(x, y) ?$$

Cuando esto ocurre se dice que $u(x, y)$ es función primitiva o **función potencial** del par de funciones $f(x, y)$, $g(x, y)$. Este último término proviene de las aplicaciones a la Física de estas ideas, lo mismo que la interpretación de la aplicación

$$D \ni (x, y) \mapsto (f(x, y), g(x, y)) \in \mathbb{R}^2$$

como un *campo vectorial* (un campo de fuerzas, con frecuencia), o sea, como una asignación a cada punto $(x, y) \in D$ de un *vector* que visualizamos con base en el punto (x, y) (Figura 5.1).

Decir que $u(x, y)$ es función potencial del par de funciones $f(x, y)$, $g(x, y)$ equivale a decir que el campo vectorial $(f(x, y), g(x, y))$ es el gradiente de

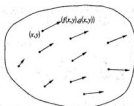


Figura 5.1

$u(x, y)$. Por eso se dice también que un campo vectorial $(f(x, y), g(x, y))$ es un **campo gradiente** si existe $u(x, y)$ tal que

$$\nabla u(x, y) \stackrel{\text{def}}{=} (u_x(x, y), u_y(x, y)) = (f(x, y), g(x, y))$$

Supuesto que $f(x, y)$ y $g(x, y)$ son de clase $C^1(D)$, si existe $u(x, y)$ tal que $u_x(x, y) = f(x, y)$, $u_y(x, y) = g(x, y)$, entonces

$$\frac{\partial f}{\partial y} = \frac{\partial}{\partial y} \left(\frac{\partial u}{\partial x} \right), \quad \frac{\partial g}{\partial x} = \frac{\partial}{\partial x} \left(\frac{\partial u}{\partial y} \right)$$

Por el teorema 5.1, estas dos derivadas han de ser iguales, de lo que se deduce que para que un campo vectorial $(f(x, y), g(x, y))$ sea un campo gradiente es **necesario** que

$$f_y(x, y) = g_x(x, y) \quad (5.2)$$

Sea, por ejemplo, el campo vectorial

$$(4x^3y^2 - \cos x, y + \sin xy)$$

Se tiene

$$f_y(x, y) = 8x^3y, \quad g_x(x, y) = y \cos xy$$

Como $f_y(x, y) \neq g_x(x, y)$, el campo vectorial no es un campo gradiente, o sea, no existe ninguna función $u(x, y)$ tal que $u_x(x, y) = 4x^3y^2 - \cos x$, $u_y(x, y) = y + \sin xy$. Vemos, por tanto, que, a diferencia de lo que ocurre para las funciones de una variable (o sea, para "campos vectoriales unidimensionales"), no todo campo vectorial $(f(x, y), g(x, y))$ tiene función primitiva o potencial.

Sea ahora el campo

$$(x^2 + y^2, 2xy + y)$$

En este caso

$$f_y(x, y) = 2y, \quad g_x(x, y) = 2y$$

es decir, sí se satisface la **condición necesaria** (5.2). ¿Existirá función potencial? Puesto que ha de cumplir que

$$\frac{\partial u}{\partial x} = x^2 + y^2 \quad (5.3)$$

ha de ser de la forma

$$u(x, y) = \frac{x^3}{3} + y^2x + C(y) \quad (5.4)$$

(resolviendo la ecuación en derivadas parciales (5.3), como vimos en el capítulo 4, integrando respecto a x y haciendo que la "constante" de integración sea, en realidad, una función arbitraria de y :

$$u(x, y) = \int (x^2 + y^2) dx + C(y) = \frac{x^3}{3} + y^2x + C(y)$$

Pero, además, $u(x, y)$ ha de satisfacer $u_y(x, y) = 2xy + y$, es decir

$$\frac{\partial}{\partial y} \left(\frac{x^3}{3} + y^2x + C(y) \right) = 2xy + C'(y) = 2xy + y$$

Por lo tanto, $C(y)$ ha de ser tal que $C'(y) = y$ o sea, de la forma

$$C(y) = \frac{y^2}{2} + C$$

con C constante arbitraria. Así pues, el campo vectorial dado es efectivamente un campo gradiente y hemos construido explícitamente una función potencial:

$$u(x, y) = \frac{x^3}{3} + y^2x + \frac{y^2}{2}$$

(no única, pues para cualquier valor de la constante C , la función $u(x, y) + C$ también lo es: en esto se tiene el mismo resultado que para las funciones de una variable, y ello era previsible pues, como sabemos, las funciones $u(x, y)$ y $u(x, y) + C$ tienen el mismo gradiente.)

El ejemplo precedente apunta a que la condición necesaria (5.2) es también **condición suficiente** para que un campo vectorial posea función potencial, sugiriendo incluso el camino de la demostración por medio de la construcción explícita de $u(x, y)$. Precisando las hipótesis, en particular en lo que respecta a la geometría del dominio D de las funciones f y g , así es en efecto:

Teorema 5.2 Supongamos que el campo vectorial $(f(x, y), g(x, y))$ está definido en un rectángulo abierto

$$R = \{(x, y); a < x < b, c < y < d\}$$

y que las derivadas parciales f_y y g_x son continuas en R . Entonces $(f(x, y), g(x, y))$ es un campo gradiente si y sólo si $f_y = g_x$.

Demostración. Ya hemos comentado la necesidad de la condición, derivada del teorema de igualdad de las derivadas cruzadas. En cuanto a la suficiencia, hemos de encontrar, dadas f y g , una función $u(x, y)$ tal que $u_x = f$, $u_y = g$. Sean (x_0, y_0) un punto fijado cualquiera y (x, y) otro punto cualquiera del rectángulo R (figura 5.2).

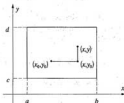


Figura 5.2

Puesto que $u(x, y)$ ha de cumplir que $u_x = f$, $u_y = g$, habrá de ser

$$u(x, y_0) = u(x_0, y_0) + \int_{x_0}^x \frac{\partial}{\partial x} u(s, y_0) ds = u(x_0, y_0) + \int_{x_0}^x f(s, y_0) ds \quad (5.5)$$

y también

$$u(x, y) = u(x, y_0) + \int_{y_0}^y \frac{\partial}{\partial y} u(x, t) dt = \int_{y_0}^y g(x, t) dt \quad (5.6)$$

de donde, sustituyendo (5.5) en (5.6),

$$u(x, y) = u(x_0, y_0) + \int_{x_0}^x f(s, y_0) ds + \int_{y_0}^y g(x, t) dt \quad (5.7)$$

(la primera integral de (5.7) se realiza a lo largo de un segmento horizontal y la segunda a lo largo de uno vertical; ambos segmentos están completamente contenidos en el rectángulo R). Así pues, si existe una función potencial $u(x, y)$ ha de venir dada por (5.7). Se trata de demostrar, por

tanto, que, efectivamente, la función $u(x, y)$ definida en (5.7) (tomando para $u(x_0, y_0)$ un valor prefijado cualquiera) verifica en R que $u_x(x, y) = f(x, y)$ y $u_y(x, y) = g(x, y)$. Por la regla de Leibniz para derivar integrales dependientes de un parámetro (apartado 4.1):

$$u_x(x, y) = f(x, y_0) + \int_{y_0}^y g_x(x, t) dt$$

que, aparentemente, no tiene nada que ver con $f(x, y)$. Pero como, por hipótesis, $g_x = f_y$, se tiene

$$u_x(x, y) = f(x, y_0) + \int_{y_0}^y f_y(x, t) dt = f(x, y_0) + f(x, y) - f(x, y_0) = f(x, y)$$

como queríamos demostrar; y de la misma manera se comprueba que $u_y(x, y) = g(x, y)$.

En la práctica, dadas f y g tales que $f_y = g_x$, se procede como en el ejemplo anterior:

- se realiza la integración respecto a la variable x

$$u(x, y) = \int f(x, y) dx + C(y) \quad (5.8)$$

con una función $C(y)$ pendiente de determinar en lugar de la constante de integración.

- el resultado se deriva respecto a y y se iguala a $g(x, y)$; la demostración del teorema garantiza que quedará para $C'(y)$ una expresión que sólo depende de y .
- se integra en y para obtener $C(y)$, se sustituye su expresión en (5.8) y queda determinada una función potencial del campo (f, g) (una infinidad de ellas, de hecho, ya que la expresión de $C(y)$ obtenida depende de una constante arbitraria).

EJEMPLO. Sea el campo

$$(y^2, 2xy + \sin y)$$

Se tiene $f_y(x, y) = 2y$, $g_x(x, y) = 2y$, por lo que se trata de un campo gradiente en cualquier rectángulo de \mathbb{R}^2 . Procediendo como se ha indicado:

$$u(x, y) = \int y^2 dx + C(y) = xy^2 + C(y)$$

$$u_y(x, y) = 2xy + C'(y) = 2xy + \sin y$$

de donde $C'(y) = \text{sen } y$, es decir $C(y) = -\cos y + C$ y

$$u(x, y) = xy^2 - \cos y + C$$

(También se puede seguir el camino inverso, integrando en primer lugar respecto a y y después respecto a x . Esto corresponde a realizar en la demostración del teorema primero la integración a lo largo del segmento vertical $[(x_0, y_0), (x_0, y)]$ y después a lo largo del segmento horizontal $[(x_0, y), (x, y)]$.)

La demostración, realmente simple, del teorema anterior se ha basado en la integración a lo largo de un segmento horizontal seguida de la integración a lo largo de uno vertical permaneciendo en el rectángulo R . De modo que la demostración también es válida para cualquier conjunto en el que sea factible esa construcción (o el camino inverso de un segmento vertical seguido de uno horizontal) para alcanzar cualquier punto (x, y) partiendo de uno fijado (x_0, y_0) , como, por ejemplo rectángulos "infinitos" —todo \mathbb{R}^2 , un semiplano, una banda, un cuadrante,...— o un círculo o un semicírculo.

EJERCICIO. Formular el correspondiente resultado para funciones de 3, y, en general, de cualquier número n de variables.

NOTA. El resultado del teorema, en su mayor generalidad, es válido para conjuntos abiertos simplemente conexos, que son aquellos conexos para los cuales el polígono delimitado por cualquier poligonal cerrada contenida en el conjunto está, también, contenido en éste (figura 5.3).

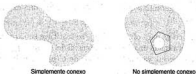


Figura 5.3

En la demostración del teorema 3 bajo esta hipótesis general es necesario utilizar integrales curvilíneas y se requieren argumentos avanzados del cálculo integral de funciones de varias variables. Si no se verifica esa restricción sobre la geometría del conjunto en el que está definido el campo vectorial, la igualdad de derivadas parciales $f_y = g_x$ ya no es suficiente, en general, para garantizar la existencia de función potencial. Un contraejemplo viene dado por

$$f(x, y) = \frac{-y}{x^2 + y^2}, \quad g(x, y) = \frac{x}{x^2 + y^2}, \quad D = \mathbb{R}^2 \setminus \{(0, 0)\} \quad (5.9)$$

Operando, se ve que si se cumple $f_y = g_x$. Supongamos que existe una función $u \in C^2(D)$ tal que $u_x = f$, $u_y = g$ para todo $(x, y) \in D$. Si integramos la ecuación $u_y = g$ en el semiplano $x > 0$ obtenemos

$$u(x, y) = \arctan \frac{y}{x} + C(x)$$

donde se entiende que, para que de verdad se trate de una función —unívoca—, se ha elegido una determinación de la función arco-tangente, digamos, la determinación principal, o sea, que $-\pi/2 < \arctan(y/x) < \pi/2$; a continuación, derivando respecto a x llegamos a $C'(x) = 0$. Los mismos cálculos valen para el semiplano $x < 0$. Por tanto, $u(x, y)$ ha de ser de la forma:

$$\begin{aligned} u(x, y) &= \arctan \frac{y}{x} + C_+ \text{ para } x > 0 \\ u(x, y) &= \arctan \frac{y}{x} + C_- \text{ para } x < 0 \end{aligned}$$

para ciertas constantes C_+ y C_- . Si la función u ha de ser continua en el semieje y positivo, tiene que verificarse $\pi/2 + C_+ = -\pi/2 + C_-$. Y si también ha de ser continua en el semieje y negativo, entonces $-\pi/2 + C_+ = \pi/2 + C_-$. Pero estas dos condiciones son incompatibles (¿por qué?) lo que implica que no puede existir función potencial del campo vectorial propuesto. (Comprobar que en el conjunto $D = \mathbb{R}^2 \setminus \{(0, y) : y < 0\}$, o sea, el plano menos el semieje y negativo, vale la demostración del teorema 3 eligiendo un (x_0, y_0) apropiado, y encontrar en él una función potencial del campo (5.9)).

Derivadas parciales de orden superior. Funciones de clase C^p

Las derivadas parciales de las derivadas segundas, si existen, se llaman **derivadas terceras** de la función f , y, en general, si se considera un conjunto de p índices $i_1, i_2, \dots, i_p \in \{1, 2, \dots, n\}$, la función

$$\frac{\partial^p f}{\partial x_{i_1} \partial x_{i_2} \dots \partial x_{i_p}} \stackrel{\text{def}}{=} \frac{\partial}{\partial x_{i_p}} \dots \left(\frac{\partial}{\partial x_{i_2}} \left(\frac{\partial f}{\partial x_{i_1}} \right) \right)$$

que se obtiene derivando p veces la función f en primer lugar respecto a x_{i_1} , después respecto a x_{i_2} y así sucesivamente hasta terminar derivando respecto a x_{i_p} , es una **derivada parcial de orden p** de f . Otras notaciones son:

$$D_{x_{i_1} x_{i_2} \dots x_{i_p}}^p f \stackrel{\text{def}}{=} D_{x_{i_p}} \dots D_{x_{i_2}} D_{x_{i_1}} f$$

$$f_{x_{i_1} x_{i_2} \dots x_{i_p}}^{(p)} \stackrel{\text{def}}{=} (((f'_{x_{i_1}})'_{x_{i_2}}) \dots)'_{x_{i_p}}$$

y también $D_{i_1 i_2 \dots i_p}^p f$, $f_{i_1 i_2 \dots i_p}^{(p)}$, $f_{x_{i_1} x_{i_2} \dots x_{i_p}}$ ó $f_{i_1 i_2 \dots i_p}$.

Como antes, $C^p(D)$ es el conjunto de todas las funciones definidas en el abierto D que tienen derivadas parciales continuas en D hasta el orden p .

Si $f \in C^p(D)$ se dice que es p veces continuamente diferenciable o que es de clase p (o de clase C^p) en D . Se conviene en que $C^0(D)$ ó $C(D)$ es el conjunto de funciones continuas en D y que $C^\infty(D)$ denota el conjunto de funciones que admiten en D derivadas parciales continuas de cualquier orden. (Por el teorema 3.2, la continuidad de las derivadas de orden p implica la continuidad de todas las derivadas de orden inferior y de la propia función.)

EJEMPLO. Para la función $f(x, y) = e^x y^2$ se tiene

$$f_{x^3}''' = D_x D_x D_x f = D_x D_x (e^x y^2) = D_x (e^x y^2) = e^x y^2$$

$$f_{x^2 y}''' = D_y D_x D_x f = D_y D_x (e^x y^2) = D_y (e^x y^2) = 2e^x y$$

$$f_{xyx}''' = D_x D_y D_x f = D_x D_y (e^x y^2) = D_x (2e^x y) = 2e^x y$$

$$f_{xy^2}''' = D_y D_y D_x f = D_y D_y (e^x y^2) = D_y (2e^x y) = 2e^x$$

$$f_{yx^2}''' = D_x D_x D_y f = D_x D_x (2e^x y) = D_y (2e^x y) = 2e^x y$$

$$f_{yxy}''' = D_y D_x D_y f = D_y D_x (2e^x y) = D_y (2e^x y) = 2e^x$$

$$f_{y^2 x}''' = D_x D_y D_y f = D_x D_y (2e^x y) = D_x (2e^x) = 2e^x$$

$$f_{y^3}''' = D_y D_y D_y f = D_y D_y (2e^x y) = D_y (2e^x) = 0$$

Como para las derivadas segundas, observamos también en este ejemplo que $f_{x^2 y}''' = f_{xyx}''' = f_{yx^2}'''$, $f_{xy^2}''' = f_{yxy}''' = f_{y^2 x}'''$, es decir, que el orden de aplicación de los operadores de derivación D_x, D_y no influye en el resultado; lo que cuenta es el número de veces que se aplica cada uno de ellos. Esta propiedad es cierta bajo hipótesis muy generales y se deduce fácilmente del teorema 5.1:

Teorema 5.3 Sea $f \in C^p(D)$. Si las listas (i_1, i_2, \dots, i_p) y (j_1, j_2, \dots, j_p) contienen el mismo número α_1 de unos, el mismo número α_2 de doses, etc., entonces

$$\frac{\partial^p f}{\partial x_{i_1} \partial x_{i_2} \cdots \partial x_{i_p}} = \frac{\partial^p f}{\partial x_{j_1} \partial x_{j_2} \cdots \partial x_{j_p}} = \frac{\partial^p f}{\partial x_1^{\alpha_1} \partial x_2^{\alpha_2} \cdots \partial x_n^{\alpha_n}} \quad (5.10)$$

En efecto, (j_1, j_2, \dots, j_p) es una permutación de los números (i_1, i_2, \dots, i_p) . Dado que una permutación cualquiera puede obtenerse por sucesivas permutaciones de dos elementos contiguos, basta demostrar el resultado para $p = 2$; pero esto es el resultado (5.1) que, como vimos, se deduce del teorema 5.1.

Demostración del teorema de igualdad de las derivadas mixtas

Sea $h \neq 0$ lo suficientemente pequeño como para que el rectángulo R_{hh} de vértices opuestos (x_0, y_0) y $(x_0 + h, y_0 + h)$ esté contenido en el entorno de (x_0, y_0) en el cual existen las derivadas f''_{xy} y f''_{yx} . Definamos

$$\begin{aligned} A(h) &\stackrel{\text{def}}{=} f(x_0 + h, y_0 + h) - f(x_0 + h, y_0) - f(x_0, y_0 + h) + f(x_0, y_0) \\ F(x) &\stackrel{\text{def}}{=} f(x, y_0 + h) - f(x, y_0), \quad G(y) \stackrel{\text{def}}{=} f(x_0 + h, y) - f(x_0, y) \end{aligned}$$

Estas funciones de una variable son derivables, y sus derivadas son

$$F'(x) = f'_x(x, y_0 + h) - f'_x(x, y_0), \quad G'(y) = f'_y(x_0 + h, y) - f'_y(x_0, y)$$

Entonces, $A(h)$ se puede escribir de las dos formas equivalentes:

$$A(h) = F(x_0 + h) - F(x_0) = G(y_0 + h) - G(y_0)$$

Aplicando el teorema del valor medio en una variable,

$$\begin{aligned} F(x_0 + h) - F(x_0) &= hF'(x_0 + \theta_1 h) = \\ &= h[f'_x(x_0 + \theta_1 h, y_0 + h) - f'_x(x_0 + \theta_1 h, y_0)] \end{aligned}$$

donde $0 < \theta_1 < 1$. Volviendo a aplicar dicho teorema, ahora a la función

$$H(y) \stackrel{\text{def}}{=} f'_x(x_0 + \theta_1 h, y),$$

cuya derivada es $H'(y) = f''_{xy}(x_0 + \theta_1 h, y)$, tendríamos

$$H(y_0 + h) - H(y_0) = hH'(y_0 + \theta_2 h)$$

donde $0 < \theta_2 < 1$. Es decir,

$$f'_x(x_0 + \theta_1 h, y_0 + h) - f'_x(x_0 + \theta_1 h, y_0) = hf''_{xy}(x_0 + \theta_1 h, y_0 + \theta_2 h),$$

y entonces

$$A(h) = h^2 f''_{xy}(x_0 + \theta_1 h, y_0 + \theta_2 h)$$

Procediendo igualmente con la segunda expresión para A , tendríamos

$$\begin{aligned} A(h) &= G(y_0 + h) - G(y_0) = hG'(y_0 + \theta_3 h) = \\ &= h[f'_y(x_0 + h, y_0 + \theta_3 h) - f'_y(x_0, y_0 + \theta_3 h)] = \\ &= h^2 f''_{yx}(x_0 + \theta_4 h, y_0 + \theta_3 h) \end{aligned}$$

donde $0 < \theta_3, \theta_4 < 1$. Por lo tanto, igualando ambas expresiones para A y dividiendo por h^2 obtenemos

$$f''_{xy}(x_0 + \theta_1 h, y_0 + \theta_2 h) = f''_{yx}(x_0 + \theta_3 h, y_0 + \theta_4 h)$$

Haciendo tender $h \rightarrow 0$, y aplicando la hipótesis de continuidad de f''_{xy} y f''_{yx} en (x_0, y_0) , llegamos a la igualdad pedida: $f''_{xy}(x_0, y_0) = f''_{yx}(x_0, y_0)$.

Hace falta imponer alguna hipótesis de regularidad para que se verifique la igualdad de las derivadas mixtas, como muestra el siguiente ejemplo.

EJEMPLO. Sea

$$f(x, y) = \begin{cases} xy \frac{x^2 - y^2}{x^2 + y^2} & \text{si } (x, y) \neq (0, 0) \\ 0 & \text{si } (x, y) = (0, 0) \end{cases}$$

Entonces

$$f'_x(x, y) = y \frac{x^4 - y^4 + 4x^2 y^2}{(x^2 + y^2)^2} \quad \text{si } (x, y) \neq (0, 0)$$

de donde $f'_x(0, y) = -y$, y $(f'_x)'_y(0, 0) = -1$. Por otra parte,

$$f'_y(x, y) = x \frac{x^4 - y^4 - 4x^2 y^2}{(x^2 + y^2)^2} \quad \text{si } (x, y) \neq (0, 0)$$

de donde $f'_y(x, 0) = x$, y $(f'_y)'_x(0, 0) = 1$. Por tanto, $f''_{xy}(0, 0) \neq f''_{yx}(0, 0)$.

Las derivadas segundas mixtas de esta función no cumplen la hipótesis del teorema. En efecto, si $(x, y) \neq (0, 0)$,

$$f''_{xy}(x, y) = f''_{yx}(x, y) = \frac{x^5 + 9x^4 y^2 - 9x^2 y^4 - y^5}{(x^2 + y^2)^3}$$

función que no tiene límite cuando $(x, y) \rightarrow (0, 0)$, como se comprueba fácilmente analizando los límites según rectas $y = mx$. Pueden relajarse algo las hipótesis del teorema, como veremos en la Nota 1 de final de capítulo.

5.2 Funciones compuestas e implícitas

Derivadas sucesivas de funciones compuestas

Sea $D \subset \mathbb{R}^2$ abierto. Supongamos que $f \in C^2(D)$ y que $u(t), v(t)$ son dos funciones de $C^2(I)$, donde I es un intervalo de \mathbb{R} , tales que $(u(t), v(t)) \in D$ para todo $t \in I$.

Como sabemos del capítulo anterior, la función compuesta

$$I \ni t \mapsto F(t) \stackrel{\text{def}}{=} f(u(t), v(t))$$

es derivable en I y su derivada viene dada por la fórmula

$$F'(t) = f'_x(u(t), v(t))u'(t) + f'_y(u(t), v(t))v'(t) \quad (5.11)$$

para todo $t \in I$.

Volvamos a derivar, y para ello observemos que $F'(t)$ está formada por una suma de productos de funciones de t por funciones compuestas de t . Por tanto, podemos aplicar de nuevo la misma regla de derivación (5.11). Para facilitar la comprensión del proceso, definamos provisionalmente

$$G(t) \stackrel{\text{def}}{=} f'_x(u(t), v(t)), \quad H(t) \stackrel{\text{def}}{=} f'_y(u(t), v(t))$$

Entonces, por (5.11),

$$\begin{aligned} G'(t) &= (f'_x)'_x(u(t), v(t))u'(t) + (f'_x)'_y(u(t), v(t))v'(t) \\ H'(t) &= (f'_y)'_x(u(t), v(t))u'(t) + (f'_y)'_y(u(t), v(t))v'(t) \end{aligned}$$

Por tanto,

$$\begin{aligned} F''(t) &= D_t [G(t)u'(t) + H(t)v'(t)] = \\ &= G'(t)u'(t) + G(t)u''(t) + H'(t)v'(t) + H(t)v''(t) = \\ &= [f''_{xx}(u(t), v(t))u'(t) + f''_{xy}(u(t), v(t))v'(t)] u'(t) + f'_x(u(t), v(t))u''(t) + \\ &\quad + [f''_{yx}(u(t), v(t))u'(t) + f''_{yy}(u(t), v(t))v'(t)] v'(t) + f'_y(u(t), v(t))v''(t) \end{aligned}$$

Reordenando términos y teniendo en cuenta que $f''_{xy} = f''_{yx}$ obtenemos la fórmula final

$$\begin{aligned} F''(t) &= f''_{xx}(u(t), v(t))u'(t)^2 + 2f''_{xy}(u(t), v(t))u'(t)v'(t) + \\ &\quad + f''_{yy}(u(t), v(t))v'(t)^2 + f'_x(u(t), v(t))u''(t) + f'_y(u(t), v(t))v''(t) \end{aligned} \quad (5.12)$$

o, simplemente,

$$F'' = f''_{xx}u'^2 + 2f''_{xy}u'v' + f''_{yy}v'^2 + f'_xu'' + f'_yv'' \quad (5.13)$$

fórmula en la que, como vemos, aparecen términos con derivadas segundas de f multiplicados por derivadas primeras de u y v , y derivadas primeras de f multiplicadas por derivadas segundas de u y v . Vemos también que $F''(t)$ es

continua, o sea, que si $f \in C^2(D)$ y $u(t), v(t) \in C^2(I)$, entonces la función compuesta

$$F(t) = f(u(t), v(t))$$

también es de clase $C^2(I)$. Reiterando el argumento se tiene que, si $f \in C^p(D)$ y $u(t), v(t) \in C^p(I)$, la aplicación repetida de la regla de la cadena implica que $t \mapsto F(t) = f(u(t), v(t))$ también es de clase $C^p(I)$ y permite calcular sus derivadas sucesivas hasta el orden p . No merece la pena intentar establecer fórmulas generales análogas a (5.12) pues son demasiado complicadas; lo haremos únicamente en el apartado 5.4 para una situación especialmente útil. Aparte de ella, lo aconsejable en cada caso concreto es derivar la expresión (5.11) utilizando la regla de la cadena las veces que sea necesario.

La fórmula análoga a (5.13) para la situación general en que $f \in C^2(D)$, siendo D un conjunto abierto de \mathbb{R}^n , $u_1(t_1, \dots, t_m), \dots, u_n(t_1, \dots, t_m)$ son funciones de clase C^2 en un abierto $D' \subset \mathbb{R}^m$ tales que $(u_1(t), \dots, u_n(t)) \in \text{Dom } f$ para todo $t = (t_1, \dots, t_m) \in D'$, y $F(t_1, \dots, t_m)$ es la función compuesta

$$D' \ni t \mapsto F(t) \stackrel{\text{def}}{=} f(u_1(t), \dots, u_n(t))$$

es

$$\frac{\partial^2 F}{\partial t_k \partial t_l} = \sum_{i,j=1}^n \frac{\partial^2 f}{\partial x_i \partial x_j} \frac{\partial u_i}{\partial t_k} \frac{\partial u_j}{\partial t_l} + \sum_{i=1}^n \frac{\partial f}{\partial x_i} \frac{\partial^2 u_i}{\partial t_k \partial t_l} \quad (5.14)$$

(Compruébese como ejercicio.)

Derivadas sucesivas de funciones implícitas

En el teorema 4.11 de la función implícita se puede probar, como veremos en el capítulo 9, que si $F(x, y)$ es de clase C^p , entonces la función $y = f(x)$ definida implícitamente por la ecuación $F(x, y) = 0$ también es de clase C^p . Aceptada la validez de este resultado, podemos aplicar, como ya se hizo para las derivadas primeras, el método general de derivación de funciones compuestas para hallar las derivadas de cualquier orden de la función implícita $y = f(x)$. Basta partir de la ecuación básica que satisface f (que supondremos definida en un intervalo I)

$$F(x, f(x)) = 0 \quad \text{para todo } x \in I$$

y derivar dos veces. Para ello, como hacíamos en el capítulo anterior, nos servimos del artificio de considerar provisionalmente una variable t que coincida

con x , de forma que se tenga $x = t$, $y = f(t)$ (es decir, $u(t) = t$, $v(t) = f(t)$), y entonces

$$D_t F(t, f(t)) = F'_x(t, f(t)) + F'_y(t, f(t))f'(t) = 0 \quad \text{para todo } x \in I$$

de donde, como vimos allí, se deduce la fórmula de la derivada primera

$$f'(t) = -\frac{F'_x(t, f(t))}{F'_y(t, f(t))} \quad \left(\text{o} \quad f'(x) = -\frac{F'_x(x, f(x))}{F'_y(x, f(x))} \right)$$

Volviendo a derivar,

$$D_t^2 F(t, f(t)) = F''_{x^2}(t, f(t)) + 2F''_{xy}(t, f(t))f'(t) + F''_{y^2}(t, f(t))f'(t)^2 + \\ + F'_x(t, f(t))0 + F'_y(t, f(t))f''(t) = 0 \quad \text{para todo } x \in I$$

de donde (retomando la variable $x = t$)

$$f''(x) = -\frac{F''_{x^2}(x, f(x)) + 2F''_{xy}(x, f(x))f'(x) + F''_{y^2}(x, f(x))f'(x)^2}{F'_y(x, f(x))} \quad (5.15)$$

que es la fórmula de la *derivada segunda de una función definida implícitamente*. Se puede sustituir en ella la fórmula anterior para f' , obteniendo la expresión equivalente

$$f''(x) = -\frac{F''_{x^2}F_y'^2 - 2F''_{xy}F'_xF'_y + F''_{y^2}F_x'^2}{F_y'^3}$$

en la que todas las derivadas están evaluadas en $(x, f(x))$. De estas fórmulas se puede sacar información acerca de la concavidad y convexidad de las funciones implícitas, máximos y mínimos locales, etc., todo ello, claro está, sin obtener una fórmula explícita para $f(x)$.

EJEMPLO 1. Consideremos la ecuación

$$F(x, y) \stackrel{\text{def}}{=} x^2 + y^2 - 1 = 0$$

que define dos funciones implícitas (continuas):

$$f_+(x) = \sqrt{1 - x^2}, \quad f_-(x) = -\sqrt{1 - x^2}$$

definidas ambas en $[-1, 1]$, y que representan las semicircunferencias superior e inferior, respectivamente, de radio 1 centradas en el origen.

Consideremos un punto cualquiera (x_0, y_0) de la curva $F = 0$, es decir, tal que

$$F(x_0, y_0) = 0$$

Como $F'_x = 2x$, $F'_y = 2y$, resulta

$$F'_x(x_0, y_0) = 2x_0, \quad F'_y(x_0, y_0) = 2y_0.$$

Por tanto, si $y_0 \neq 0$, el teorema de la función implícita nos garantiza la existencia de una función $f : (x_0 - \delta, x_0 + \delta) \rightarrow (y_0 + \varepsilon, y_0 + \varepsilon)$ que satisface $x^2 + f(x)^2 - 1 = 0$ y $f(x_0) = y_0$. Obviamente, si $y_0 > 0$, $f(x) = f_+(x)$, y si $y_0 < 0$, $f(x) = f_-(x)$, así que podemos garantizar que f está definida en $[-1, 1]$.

La derivada $f'(x)$ satisface

$$f'(x) = -\frac{x}{f(x)} \quad \text{para todo } x \in (-1, 1)$$

tanto en un caso como en el otro. En particular, $f'(0) = 0$.

La derivada segunda se puede obtener directamente:

$$f''(x) = -\frac{f(x) - xf'(x)}{f(x)^2}$$

o bien a partir de la fórmula general

$$f''(x) = -\frac{F''_{x^2}(x, f(x)) + 2F''_{xy}(x, f(x))f'(x) + F''_{y^2}(x, f(x))f'(x)^2}{F'_y(x, f(x))}$$

obteniendo

$$f''(x) = -\frac{2 + 2f'(x)^2}{2f(x)}$$

que conducen al mismo resultado que la segunda fórmula general

$$f''(x) = -\frac{F''_{x^2}F_y'^2 - 2F''_{xy}F'_xF'_y + F''_{y^2}F_x'^2}{F_y'^3}$$

es decir:

$$f''(x) = -\frac{2(2f(x))^2 + 2(2x)^2}{(2f(x))^3} = -\frac{f(x)^2 + x^2}{f(x)^3}$$

como se puede comprobar fácilmente. Particularizando en (x_0, y_0) ,

$$f''(x_0) = -\frac{f(x_0)^2 + x_0^2}{f(x_0)^3} = -\frac{y_0^2 + x_0^2}{y_0^3}$$

y, concretamente,

$$f''(0) = -\frac{1}{y_0}$$

En consecuencia: Si $y_0 > 0$, f (que coincide con f_+) alcanza un *máximo local* para $x = 0$, mientras que si $y_0 < 0$, f (que ahora coincidiría con f_-) alcanza un *mínimo local* para $x = 0$. Ambas conclusiones se corresponden, evidentemente, con la interpretación geométrica de estas funciones implícitas como semicircunferencias superior (para el máximo) e inferior (para el mínimo) de la circunferencia unidad. Obsérvese también como de la expresión para f'' se deduce que f_+ es cóncava y f_- convexa en $[-1, 1]$.

EJEMPLO 2. Consideremos un ejemplo parecido al número 7 del apartado 4.2:

$$F(x, y) = x - \ln x + y - 2 \ln y - 2$$

Sea $(x_0, y_0) = (1, 1)$. Entonces:

- $F(1, 1) = 0$
- $F'_x(x, y) = 1 - \frac{1}{x}$, $F'_y(x, y) = 1 - \frac{2}{y}$, de donde $F'_y(1, 1) \neq 0$, y $F(x, y) = 0$ define implícitamente una función $f : (1 - \delta, 1 + \delta) \rightarrow (1 - \varepsilon, 1 + \varepsilon)$ tal que

$$\begin{aligned} f(1) &= 1 \\ F(x, f(x)) &= 0 \quad \text{para todo } x \in (1 - \delta, 1 + \delta) \\ f'(x) &= -\frac{1 - \frac{1}{x}}{1 - \frac{2}{f(x)}} \quad \text{para todo } x \in (1 - \delta, 1 + \delta) \end{aligned}$$

En particular, $f'(1) = 0$.

- $F''_{xx}(x, y) = \frac{1}{x^2}$, $F''_{xy}(x, y) = 0$, $F''_{yy}(x, y) = \frac{2}{y^2}$. Por tanto,

$$\begin{aligned} f''(x) &= -\frac{F''_{xx}(x, f(x)) + 2F''_{xy}(x, f(x))f'(x) + F''_{yy}(x, f(x))f'(x)^2}{F'_y(x, f(x))} = \\ &= -\frac{\frac{1}{x^2} + \frac{2}{f(x)^2}f'(x)^2}{1 - \frac{2}{f(x)}} \end{aligned}$$

o bien

$$\begin{aligned} f''(x) &= -\frac{F_{x^2}'' F_y'^2 - 2F_{xy}'' F_x' F_y' + F_{y^2}'' F_x'^2}{F_y'^3} = \\ &= -\frac{\frac{1}{x^2} \left(1 - \frac{2}{f(x)}\right)^2 + \frac{2}{f(x)^2} \left(1 - \frac{1}{x}\right)}{\left(1 - \frac{2}{f(x)}\right)^3}. \end{aligned}$$

En cualquier caso —mejor a partir de la primera de las fórmulas, pues el saber que $f'(1) = 0$ simplifica los cálculos—, observamos que

$$f''(1) = -\frac{1}{-1} = 1 > 0$$

Por tanto, la función implícita f alcanza un *mínimo local* para $x = 1$.

Como dijimos en el apartado 4.2, los cálculos suelen hacerse en términos mucho más informales que los aquí presentados. Por ejemplo, de

$$x - \ln x + y - 2 \ln y - 2 = 0$$

concluimos (haciendo mentalmente $y = f(x)$ y escribiendo $f' = y'$)

$$1 - \frac{1}{x} + y' - \frac{2y'}{y} = 1 - \frac{1}{x} + \left(1 - \frac{2}{y}\right)y' = 0 \quad (5.16)$$

de donde

$$y' = -\frac{1 - \frac{1}{x}}{1 - \frac{2}{y}} \quad (5.17)$$

Para calcular la derivada segunda se puede usar directamente esta fórmula, obteniendo:

$$y'' = -\frac{\frac{1}{x^2} \left(1 - \frac{2}{y}\right) - \left(1 - \frac{1}{x}\right) \frac{2y'}{y^2}}{\left(1 - \frac{2}{y}\right)^2}$$

y combinarla con la obtenida para y' , o bien derivar directamente en (5.16):

$$\frac{1}{x^2} + \frac{2y'}{y^2} y' + \left(1 - \frac{2}{y}\right) y'' = 0$$

EJERCICIO. Dada la ecuación

$$y^5 - y^2 + 2xy - x^2 = 0$$

comprobar que

$$y' = -\frac{2y - 2x}{5y^4 - 2y + 2x}, \quad y'' = -\frac{(20y^3 - 2)(y')^2 + 4y' - 2}{5y^4 - 2y + 2x}$$

5.3 Fórmula de Taylor I

Mediante el desarrollo de Taylor se aproxima una función en el entorno de un punto por polinomios de orden creciente, mejorando la aproximación lineal, o de *primer orden*, que proporciona la diferencial. Nos centraremos aquí en el desarrollo de Taylor de *segundo orden*, la situación más simple y que tiene las aplicaciones más inmediatas, en particular a los problemas de optimización que veremos en el capítulo siguiente. Los desarrollos de Taylor de orden superior a dos, de utilidad menos inmediata y que requieren expresiones inevitablemente complicadas, se tratarán con todo detalle en el apartado 5.4 que sigue; su lectura no será necesaria para el resto del libro.

Comencemos repasando la fórmula de Taylor para funciones de una variable, pues, como decíamos, en ella se apoya directamente el correspondiente resultado para funciones de varias variables. Sea $y = f(x)$ una función de una variable de clase C^2 en un intervalo I . Dado $x_0 \in I$, recordemos que la diferencial representa la aproximación lineal del incremento:

$$\Delta f(x_0; h) \stackrel{\text{def}}{=} f(x_0 + h) - f(x_0) = df(x_0; h) + R_1(h) = f'(x_0)h + R_1(h) \quad (5.18)$$

donde $R_1(h) = o(h)$, es decir

$$\lim_{h \rightarrow 0} \frac{R_1(h)}{h} = 0 \quad (5.19)$$

(Recuérdese que, si existe $f'(x_0)$, entonces vale la fórmula (5.18), con la propiedad (5.19), por la definición de derivada; y si $\Delta f(x_0; h)$ admite una expresión de la forma $ah + o(h)$, entonces ha de ser necesariamente $a = f'(x_0)$.) Escribiendo (5.18) en la forma

$$f(x_0 + h) = f(x_0) + f'(x_0)h + R_1(h) \quad (5.20)$$

vemos que los valores de f en puntos próximos a x_0 se aproximan por los del polinomio de primer grado $f(x_0) + f'(x_0)h$ con un error, dependiente de h ,

que converge a cero "más rápidamente" que la propia distancia $|h|$ de dichos puntos a x_0 , que es lo que quiere decir la propiedad (5.19).

¿Cómo podríamos mejorar esa aproximación? Teniendo en cuenta los comentarios que hicimos en el tercer capítulo acerca de los órdenes de magnitud, la idea más lógica sería añadir a la diferencial un término *cuadrático*:

$$f(x_0 + h) - f(x_0) = f'(x_0)h + Bh^2 + R_2(h) \quad (5.21)$$

de forma que ahora, con una elección adecuada de B , tuviésemos $R_2(h) = o(h^k)$, con el mayor k posible. La elección $B = 0$ nos da $k = 1$, que es lo mejor que se puede lograr con la aproximación lineal. ¿Podremos conseguir $k = 2$? Nos planteamos, pues, si existe algún B para el cual se tenga

$$\lim_{h \rightarrow 0} \frac{f(x_0 + h) - f(x_0) - f'(x_0)h - Bh^2}{h^2} = 0$$

o sea,

$$\lim_{h \rightarrow 0} \frac{f(x_0 + h) - f(x_0) - f'(x_0)h}{h^2} - B = 0$$

El problema se reduce a estudiar la existencia de este límite, y a calcularlo. Se trata de un límite indeterminado de la forma $0/0$, al cual se puede intentar aplicar la regla de l'Hôpital: si existe el límite del cociente de las derivadas (como funciones de h), entonces éste coincide con el límite buscado. Pero este límite es

$$\lim_{h \rightarrow 0} \frac{f'(x_0 + h) - f'(x_0)}{2h} = \frac{1}{2}f''(x_0)$$

por la propia definición de la derivada segunda. Por tanto, la única elección posible es $B = f''(x_0)/2$, y recíprocamente, con esta elección, la expresión cuadrática $f'(x_0)h + \frac{1}{2}f''(x_0)h^2$ aproxima al incremento con un error $o(h^2)$:

$$f(x_0 + h) - f(x_0) = f'(x_0)h + \frac{1}{2}f''(x_0)h^2 + o(h^2) \quad (5.22)$$

Esta es la **fórmula de Taylor de segundo orden** (la de **primer orden** es la fórmula (5.18) de la definición de diferencial). Escribiéndola en la forma

$$f(x_0 + h) = f(x_0) + f'(x_0)h + \frac{1}{2}f''(x_0)h^2 + o(h^2) \quad (5.23)$$

vemos como los valores de f en puntos próximos a x_0 se aproximan por los valores del polinomio de segundo grado

$$P_2(x_0; h) = f(x_0) + f'(x_0)h + \frac{1}{2}f''(x_0)h^2 \quad (5.24)$$

y que esta aproximación es mejor que la aproximación lineal (5.20) pues el error que se comete es un infinitésimo de mayor orden que el cuadrado de la distancia de dichos puntos a x_0 , $|h|^2$, que, a su vez, lo es de mayor orden que la distancia $|h|$ cuando $|h| \rightarrow 0$: $|h|^2/|h| = |h| \rightarrow 0$. El polinomio $P_2(x_0; h)$ definido en (5.24) es el **polinomio de Taylor de segundo grado** de f en x_0 . La cantidad

$$R_2(x_0; h) = f(x_0 + h) - P_2(x_0; h) \quad (5.25)$$

es el **resto o error de segundo orden** que se comete al sustituir $f(x_0 + h)$ por $P_2(x_0; h)$. La expresión

$$d^2 f(x_0; h) \stackrel{\text{def}}{=} f''(x_0)h^2 \quad (5.26)$$

es la **diferencial segunda de f en x_0** . El nombre y su notación $d^2 f$ se debe a que viene a ser algo así como "la diferencial de la diferencial" en el sentido siguiente: Dada la fórmula que define la diferencial (primera) de f :

$$df(x; h) \stackrel{\text{def}}{=} f'(x)h$$

con x recorriendo I , fijemos h , y calculemos la diferencial de df como función de x solamente. Para ello, llamemos $g(x) \stackrel{\text{def}}{=} f'(x)h$, cuya derivada vale $g'(x) = f''(x)h$ (no olvidemos que h está fijado) con lo cual $dg(x, h) = g'(x)h = (f''(x)h)h = f''(x)h^2$. Como la diferencial primera, es una función de dos variables: x y h . Con la notación (5.26), la fórmula de Taylor de segundo orden se escribe en la "forma diferencial":

$$f(x_0 + h) - f(x_0) = df(x_0; h) + \frac{1}{2}d^2 f(x_0; h) + o(h^2) \quad (5.27)$$

La fórmula de segundo orden (5.22) se puede también obtener a partir de la de primer orden (5.18) que expresa la diferenciabilidad de f cuando se supone, como estamos haciendo aquí, que $f \in C^2(I)$. Esta hipótesis extra nos permite obtener una expresión en términos de f'' para el resto $R_1(h)$ de la manera que sigue. Recordemos que el **teorema de Cauchy** para funciones de una variable, que es una generalización del teorema del valor medio, dice que si $f(x)$ y $g(x)$ son continuas en $[a, b]$, derivables en (a, b) y $g'(x) \neq 0$ para $x \in (a, b)$, entonces existe un punto $\xi \in (a, b)$ tal que

$$\frac{f(b) - f(a)}{g(b) - g(a)} = \frac{f'(\xi)}{g'(\xi)} \quad (5.28)$$

(Nótese que $g(b) - g(a) \neq 0$ gracias a la hipótesis $g'(x) \neq 0$ en (a, b) —¿por qué?—. Este resultado se obtiene fácilmente a partir del teorema de Rolle: se

considera la función auxiliar

$$F(x) = f(x) - f(a) - \frac{f(b) - f(a)}{g(b) - g(a)}[g(x) - g(a)]$$

Por hipótesis, F es continua en $[a, b]$, derivable en (a, b) y verifica $F(a) = 0$, $F(b) = 0$, con lo que, por el teorema de Rolle, existe $\xi \in (a, b)$ tal que $F'(\xi) = 0$; sustituyendo, se tiene (5.28). Del teorema de Cauchy se deduce, con $g(x) = x$, el teorema del valor medio de Lagrange. Obsérvese que la fórmula (5.28) vale también cuando $a > b$.

Pues bien, partiendo, para un x_0 fijado, de

$$R_1(t) = f(x_0 + t) - f(x_0) - f'(x_0)t, \quad t \in [0, h]$$

y aplicando el teorema de Cauchy a las funciones $R_1(t)$ y $\varphi(t) = t^2$ en el intervalo $t \in [0, h]$ se tendrá, contando con que $R_1(0) = R'_1(0) = 0$, $R'_1(t) = f''(x_0 + t)$, $\varphi(0) = \varphi'(0) = 0$, $\varphi''(0) = 2$,

$$\begin{aligned} \frac{R_1(h)}{\varphi(h)} &= \frac{R_1(h) - R_1(0)}{\varphi(h) - \varphi(0)} = \frac{R'_1(\theta_1 h)}{\varphi'(\theta_1 h)} = \\ &= \frac{R'_1(\theta_1 h) - R'_1(0)}{\varphi'(\theta_1 h) - \varphi'(0)} = \frac{R''_1(\theta_2 \theta_1 h)}{\varphi''(\theta_2 \theta_1 h)} = \\ &= \frac{R''_1(\theta h)}{2} = \frac{1}{2} f''(x_0 + \theta h) \end{aligned}$$

con $\theta_1, \theta_2 \in (0, 1)$ y $\theta = \theta_2 \theta_1 \in (0, 1)$, es decir

$$R_1(h) = \frac{1}{2} f''(x_0 + \theta h) h^2 \quad (5.29)$$

Así pues, si $f \in C^2(I)$ se tiene para cada $x_0 \in I$

$$f(x_0 + h) - f(x_0) = f'(x_0)h + \frac{1}{2} f''(x_0 + \theta h) h^2 \quad (5.30)$$

para un cierto $\theta \in (0, 1)$ (h puede ser positivo o negativo). Esta es la fórmula de Taylor de primer orden con resto de Lagrange. Constituye, de hecho, una formulación de la diferenciabilidad de f con la precisión suplementaria de una expresión "cuasi-explicita" — θ no se conoce exactamente— para el resto $R_1(h)$ (obsérvese como efectivamente

$$\frac{R_1(h)}{h} = \frac{1}{2} f''(x_0 + \theta h) \frac{h^2}{h} \xrightarrow{h \rightarrow 0} 0;$$

en la nota 2 de final de capítulo veremos otras expresiones del resto para desarrollos de Taylor de cualquier orden).

A partir de (5.30) se obtiene fácilmente la fórmula de Taylor (5.23). En efecto, escribiendo a partir de (5.30)

$$f(x_0 + h) - f(x_0) = f'(x_0)h + \frac{1}{2}f''(x_0)h^2 + \frac{1}{2}[f''(x_0 + \theta h) - f''(x_0)]h^2$$

lo único que hay que probar es que

$$\frac{\frac{1}{2}[f''(x_0 + \theta h) - f''(x_0)]h^2}{h^2} = \frac{1}{2}[f''(x_0 + \theta h) - f''(x_0)] \rightarrow 0$$

cuando $h \rightarrow 0$; pero esto es evidente por la continuidad de f'' en x_0 .

La extensión de estas ideas a funciones de dos y más variables se obtiene fácilmente a partir de los resultados para una variable aplicando la regla de la cadena para calcular derivadas sucesivas de funciones compuestas tal como se ha expuesto en el apartado anterior.

Sea $f \in C^2(D)$, con D un conjunto abierto de \mathbb{R}^2 . Sea $(x_0, y_0) \in D$ y sean (h, k) tales que el segmento que une (x_0, y_0) con $(x_0 + h, y_0 + k)$ está contenido en D . Consideremos la función compuesta

$$F(t) = f(x_0 + th, y_0 + tk), \quad 0 \leq t \leq 1 \quad (5.31)$$

dada por la evaluación de f a lo largo de dicho segmento. (Es la misma función que utilizábamos en el apartado 4.1 para demostrar el teorema del valor medio para funciones de varias variables). Por (5.30):

$$F(1) = F(0) + F'(0) + \frac{1}{2}F''(\theta), \quad 0 < \theta < 1 \quad (5.32)$$

Se tiene

$$F(0) = f(x_0, y_0), \quad F(1) = f(x_0 + h, y_0 + k)$$

Por la regla de la cadena

$$F'(t) = hf'_x(x_0 + th, y_0 + tk) + kf'_y(x_0 + th, y_0 + tk) \quad (5.33)$$

y, por tanto,

$$F'(0) = hf'_x(x_0, y_0) + kf'_y(x_0, y_0)$$

Finalmente, necesitamos calcular $F''(t)$ para expresar (5.32) en términos de la función $f(x, y)$. Por la fórmula (5.12) del apartado anterior, teniendo en cuenta que aquí $u''(t) = v''(t) = 0$, se tiene

$$\begin{aligned} F''(t) &= h^2 f_{xx}(x_0 + th, y_0 + tk) + 2hk f_{xy}(x_0 + th, y_0 + tk) + \\ &\quad + k^2 f_{yy}(x_0 + th, y_0 + tk) \end{aligned} \quad (5.34)$$

En consecuencia, (5.32) es equivalente a

$$f(x_0 + h, y_0 + k) = f(x_0, y_0) + hf'_x(x_0, y_0) + kf'_y(x_0, y_0) + \frac{1}{2} [h^2 f''_{xx}(X, Y) + 2hk f''_{xy}(X, Y) + k^2 f''_{yy}(X, Y)] \quad (5.35)$$

donde $(X, Y) = (x_0 + \theta h, y_0 + \theta k)$, $0 < \theta < 1$.

La fórmula (5.35) es el desarrollo de Taylor de primer orden con resto de Lagrange; de hecho, constituye, como decíamos antes para las funciones de una variable, una expresión de la diferenciabilidad de f en (x_0, y_0) :

$$f(x_0 + h, y_0 + k) = f(x_0, y_0) + hf'_x(x_0, y_0) + kf'_y(x_0, y_0) + g(h, k)$$

con la precisión suplementaria de una expresión para el resto $g(h, k)$; gracias a (5.35) se puede establecer en qué medida la función f se puede aproximar por un polinomio de segundo grado:

Teorema 5.4 Sea $D \subset \mathbb{R}^2$ abierto y supongamos que $f : D \rightarrow \mathbb{R}$ tiene derivadas parciales continuas hasta el orden 2. Sea $(x_0, y_0) \in D$ y sean (h, k) tales que el segmento de extremos (x_0, y_0) y $(x_0 + h, y_0 + k)$ está contenido en D . Entonces

$$f(x_0 + h, y_0 + k) = f(x_0, y_0) + hf'_x(x_0, y_0) + kf'_y(x_0, y_0) + \frac{1}{2} [h^2 f''_{xx}(x_0, y_0) + 2hk f''_{xy}(x_0, y_0) + k^2 f''_{yy}(x_0, y_0)] + R_2(h, k) \quad (5.36)$$

con $R_2(h, k)$ tal que

$$\lim_{(h,k) \rightarrow (0,0)} \frac{R_2(h, k)}{h^2 + k^2} = 0 \quad (5.37)$$

Demostración. Se trata de probar que

$$\lim_{(h,k) \rightarrow (0,0)} \frac{f(x_0 + h, y_0 + k) - P_2(h, k)}{h^2 + k^2} = 0 \quad (5.38)$$

donde $P_2(h, k)$ es el polinomio de segundo grado en h, k

$$P_2(h, k) = f(x_0, y_0) + hf'_x(x_0, y_0) + kf'_y(x_0, y_0) + \frac{1}{2} [h^2 f''_{xx}(x_0, y_0) + 2hk f''_{xy}(x_0, y_0) + k^2 f''_{yy}(x_0, y_0)] \quad (5.39)$$

Por (5.35)

$$\begin{aligned} & \frac{f(x_0 + h, y_0 + k) - P_2(h, k)}{h^2 + k^2} = \\ &= \frac{h^2}{2(h^2 + k^2)} [f''_{xx}(x_0 + \theta h, y_0 + \theta k) - f''_{xx}(x_0, y_0)] + \\ &+ \frac{2hk}{2(h^2 + k^2)} [f''_{xy}(x_0 + \theta h, y_0 + \theta k) - f''_{xy}(x_0, y_0)] + \\ &+ \frac{k^2}{2(h^2 + k^2)} [f''_{yy}(x_0 + \theta h, y_0 + \theta k) - f''_{yy}(x_0, y_0)] \end{aligned}$$

Puesto que

$$0 < \frac{h^2}{2(h^2 + k^2)} \leq \frac{1}{2}, 0 < \frac{2|hk|}{2(h^2 + k^2)} \leq \frac{h^2 + k^2}{2(h^2 + k^2)} \leq \frac{1}{2}, 0 < \frac{k^2}{2(h^2 + k^2)} \leq \frac{1}{2}$$

para $(h, k) \neq (0, 0)$, y los términos entre corchetes convergen a 0 cuando $(h, k) \rightarrow (0, 0)$ por la continuidad de las derivadas segundas, queda demostrado (5.38).

La expresión (5.36) es el desarrollo de Taylor de segundo orden de f en (x_0, y_0) . Valen aquí los mismos comentarios que hacíamos en el caso de una variable: la diferenciabilidad de f en (x_0, y_0) significa que

$$\lim_{(h,k) \rightarrow (0,0)} \frac{f(x_0 + h, y_0 + k) - P_1(h, k)}{\sqrt{h^2 + k^2}} = 0 \quad (5.40)$$

donde

$$\begin{aligned} P_1(h, k) &= f(x_0, y_0) + hf'_x(x_0, y_0) + kf'_y(x_0, y_0) \stackrel{\text{def}}{=} \\ &= f(x_0, y_0) + df(x_0, y_0; h, k) \end{aligned} \quad (5.41)$$

o sea, los valores de f en puntos próximos a (x_0, y_0) se aproximan por un polinomio de primer grado (función lineal) con un error que es un infinitésimo de mayor orden que la distancia $\sqrt{h^2 + k^2}$ de dichos puntos a (x_0, y_0) . El resultado (5.37) del teorema anterior significa que el polinomio de segundo grado $P_2(h, k)$ definido en (5.39) aproxima mejor los valores de f pues el error que se comete es un infinitésimo de mayor orden que el cuadrado de aquella distancia, que, a su vez, lo es de mayor orden que $\sqrt{h^2 + k^2}$ cuando $(h, k) \rightarrow (0, 0)$. $P_2(h, k)$ es el polinomio de Taylor de segundo grado de f en (x_0, y_0) . La cantidad $R_2(h, k) = f(x_0 + h, y_0 + k) - P_2(h, k)$ es el resto o error de segundo orden que se comete al sustituir $f(x_0 + h, y_0 + k)$ por $P_2(h, k)$. Descrito únicamente por la propiedad (5.37), $R_2(h, k) = o(h^2 + k^2)$,

se dice que el resto está en forma asintótica o de Peano. (Con mayor regularidad sobre la función f , podremos expresarlo, como se demuestra en el apartado siguiente, de una forma más precisa que implica ese comportamiento asintótico; concretamente, si f es de clase C^3 , $R_2(h, k)$ se puede expresar en términos de las derivadas terceras de f evaluadas en un punto intermedio $(x_0 + \theta h, y_0 + \theta k)$.)

Así como, siendo $F(t) = f(x_0 + th, y_0 + tk)$, se tiene

$$df(x_0, y_0; h, k) = F'(0) = hf'_x(x_0, y_0) + kf'_y(x_0, y_0) \quad (5.42)$$

se define la **diferencial segunda de f en (x_0, y_0)** como la función de $(x_0, y_0; h, k)$ dada por

$$d^2f(x_0, y_0; h, k) \stackrel{\text{def}}{=} F''(0) = h^2 f''_{xx}(x_0, y_0) + 2hk f''_{xy}(x_0, y_0) + k^2 f''_{yy}(x_0, y_0) \quad (5.43)$$

Para (x_0, y_0) fijo, es un polinomio homogéneo de segundo grado en (h, k) , o sea, una *forma cuadrática*, que tiene la particularidad de poderse expresar cómodamente en *forma matricial*:

$$d^2f(x_0, y_0; h, k) = (h, k) \begin{pmatrix} f''_{xx}(x_0, y_0) & f''_{xy}(x_0, y_0) \\ f''_{xy}(x_0, y_0) & f''_{yy}(x_0, y_0) \end{pmatrix} \begin{pmatrix} h \\ k \end{pmatrix}$$

(como se comprueba inmediatamente multiplicando las matrices indicadas). La matriz

$$\begin{pmatrix} f''_{xx}(x_0, y_0) & f''_{xy}(x_0, y_0) \\ f''_{xy}(x_0, y_0) & f''_{yy}(x_0, y_0) \end{pmatrix}$$

recibe el nombre de *matriz hessiana de f en (x_0, y_0)* , y es simétrica. La representaremos por $D^2f(x_0, y_0)$. Su determinante se llama *hessiano de f en (x_0, y_0)* , y lo representaremos por $Hf(x_0, y_0)$.

Con esta definición de diferencial segunda, el polinomio de Taylor de segundo grado de f toma la forma

$$P_2(h, k) = f(x_0, y_0) + df(x_0, y_0; h, k) + \frac{1}{2}d^2f(x_0, y_0; h, k) \quad (5.44)$$

y la fórmula de aproximación (5.36):

$$f(x_0 + h) = f(x_0, y_0) + df(x_0, y_0; h, k) + \frac{1}{2}d^2f(x_0, y_0; h, k) + o(h^2 + k^2) \quad (5.45)$$

es formalmente la misma que en el caso de una variable.

Es frecuente escribir la fórmula (5.39) que define el polinomio de Taylor de segundo grado así como la fórmula de aproximación (5.36) destacando el papel del punto $(x, y) = (x_0 + h, y_0 + k)$ al que se llega tras aplicar el incremento (h, k) :

$$\begin{aligned} P_2(x, y) \stackrel{\text{def}}{=} & f(x_0, y_0) + f'_x(x_0, y_0)(x - x_0) + f'_y(x_0, y_0)(y - y_0) + \\ & + \frac{1}{2} [f''_{xx}(x_0, y_0)(x - x_0)^2 + 2f''_{xy}(x_0, y_0)(x - x_0)(y - y_0) + \\ & + f''_{yy}(x_0, y_0)(y - y_0)^2] \end{aligned} \quad (5.46)$$

$$\begin{aligned} f(x, y) = & f(x_0, y_0) + f'_x(x_0, y_0)(x - x_0) + f'_y(x_0, y_0)(y - y_0) + \\ & + \frac{1}{2} [f''_{xx}(x_0, y_0)(x - x_0)^2 + 2f''_{xy}(x_0, y_0)(x - x_0)(y - y_0) + \\ & + f''_{yy}(x_0, y_0)(y - y_0)^2] + o((x - x_0)^2 + (y - y_0)^2) \end{aligned} \quad (5.47)$$

Recordemos la interpretación geométrica de la diferenciabilidad: la gráfica de la función $z = f(x, y)$, una superficie en \mathbb{R}^3 , se puede aproximar en las cercanías de (x_0, y_0) por el plano tangente

$$z = f(x_0, y_0) + f'_x(x_0, y_0)(x - x_0) + f'_y(x_0, y_0)(y - y_0) \quad (5.48)$$

con error $o(\sqrt{(x - x_0)^2 + (y - y_0)^2})$. Ahora, la fórmula (5.47) nos dice, desde el punto de vista geométrico, que dicha gráfica se puede aproximar mejor por un paraboloide de ecuación $z = P_2(x, y)$ que tiene en (x_0, y_0) el mismo plano tangente que $z = f(x, y)$ (¿por qué?). Se pierde en simplicidad de la función aproximante pero se gana en orden de aproximación.

EJEMPLOS.

1. Si $f(x, y) = e^x \sin y$, se tiene

$$f_x(x, y) = e^x \sin y, \quad f_y(x, y) = e^x \cos y$$

$$f_{xx}(x, y) = e^x \sin y$$

$$f_{xy}(x, y) = e^x \cos y = f_{yx}(x, y)$$

$$f_{yy}(x, y) = -e^x \sin y$$

y en consecuencia

$$\begin{aligned} e^x \sin y = & e^{x_0} \sin y_0 + (e^{x_0} \sin y_0)(x - x_0) + (e^{x_0} \cos y_0)(y - y_0) + \\ & + \frac{1}{2} [(e^{x_0} \sin y_0)(x - x_0)^2 + 2(e^{x_0} \cos y_0)(x - x_0)(y - y_0) - \\ & - (e^{x_0} \sin y_0)(y - y_0)^2] + o((x - x_0)^2 + (y - y_0)^2) \end{aligned}$$

Si, por ejemplo, $(x_0, y_0) = (0, 0)$, se tiene

$$e^x \operatorname{sen} y = y + xy + o(x^2 + y^2) = y(1 + x) + o(x^2 + y^2)$$

2. Obtengamos el polinomio de Taylor de segundo grado en el punto $(1, 0)$ de la función

$$f(x, y) = x^y$$

Para ello, calculemos las derivadas parciales de primer y segundo orden:

$$D_x(x^y) = x^{y-1}y$$

$$D_y(x^y) = x^y \ln x$$

$$D_{xx}(x^y) = x^{y-2}y^2 - x^{y-2}y = (y-1)x^{y-2}y$$

$$D_{xy}(x^y) = x^{y-1}(\ln x)y + x^{y-1}$$

$$D_{yy}(x^y) = x^y(\ln x)^2$$

Evaluando la función y sus derivadas en $(1, 0)$, obtenemos:

$$f(1, 0) = 1^0 = 1$$

$$f'_x(1, 0) = 1^{-1}0 = 0$$

$$f'_y(1, 0) = 1^0 \ln 1 = 0$$

$$f''_{xx}(1, 0) = (-1)1^{-2}0 = 0$$

$$f''_{xy}(1, 0) = 1^{-1}(\ln 1)0 + 1^{-1} = 1$$

$$f''_{yy}(1, 0) = 1^0(\ln 1)^2 = 0$$

Por tanto,

$$\begin{aligned} P_2(x, y) &= 1 + 0(x-1) + 0y + \frac{1}{2}(0(x-1)^2 + 2 \cdot 1(x-1)y + 0y^2) = \\ &= (x-1)y \end{aligned}$$

o, llamando $h = x - 1$, $k = y - 0 = y$,

$$\tilde{P}_2(h, k) = P_2(1+h, k) = hk$$

donde hemos denotado \tilde{P}_2 al polinomio de Taylor como función de las variables h y k .

Las correspondientes fórmulas para funciones $f(x_1, \dots, x_n)$ de n variables se establecen de la misma manera:

Si $f \in C^2(D)$, $D \subset \mathbb{R}^n$ abierto, $x_0 = (x_1^0, \dots, x_n^0) \in D$ y $h = (h_1, \dots, h_n)$ son tales que el segmento que une x_0 con $x_0 + h$ está contenido en D , entonces se tiene:

1. (Fórmula de Taylor de primer orden con resto de Lagrange)

$$f(x_0 + h) = f(x_0) + [f'_{x_1}(x_0)h_1 + \dots + f'_{x_n}(x_0)h_n] + \\ + \frac{1}{2} \sum_{i,j=1}^n f''_{x_i x_j}(x_0 + \theta h) h_i h_j, \quad (5.49)$$

para algún $0 < \theta < 1$.

2. (Fórmula de Taylor de segundo orden con resto de Peano)

$$f(x_0 + h) = f(x_0) + [f'_{x_1}(x_0)h_1 + \dots + f'_{x_n}(x_0)h_n] + \\ + \frac{1}{2} \sum_{i,j=1}^n f''_{x_i x_j}(x_0) h_i h_j + o(|h|^2) \quad (5.50)$$

3.

$$f(x) = f(x_0) + [f'_{x_1}(x_0)(x_1 - x_1^0) + \dots + f'_{x_n}(x_0)(x_n - x_n^0)] + \\ + \frac{1}{2} \sum_{i,j=1}^n f''_{x_i x_j}(x_0)(x_i - x_i^0)(x_j - x_j^0) + o(|x - x_0|^2) \quad (5.51)$$

La función auxiliar para obtener estos resultados es

$$F(t) = f(x_0 + th) = f(x_1^0 + th_1, \dots, x_n^0 + th_n)$$

La diferencial segunda de f en x_0 es

$$d^2 f(x_0; h) \stackrel{\text{def}}{=} F''(0) = \sum_{i,j=1}^n f''_{x_i x_j}(x_0) h_i h_j \quad (5.52)$$

o en forma matricial

$$d^2 f(x_0; h) = (h_1, \dots, h_n) \begin{pmatrix} f''_{x_1 x_1}(x_0) & \dots & f''_{x_1 x_n}(x_0) \\ \vdots & \dots & \vdots \\ f''_{x_n x_1}(x_0) & \dots & f''_{x_n x_n}(x_0) \end{pmatrix} \begin{pmatrix} h_1 \\ \vdots \\ h_n \end{pmatrix} \quad (5.53)$$

La matriz

$$\begin{pmatrix} f''_{x_1 x_1}(x_0) & \dots & f''_{x_1 x_n}(x_0) \\ \vdots & \dots & \vdots \\ f''_{x_n x_1}(x_0) & \dots & f''_{x_n x_n}(x_0) \end{pmatrix}$$

es la *matriz hessiana* de f en x_0 ; es simétrica por la igualdad de las derivadas cruzadas y se representa por $D^2 f(x_0)$. Su determinante es el *hessiano* de f en x_0 y se representa por $Hf(x_0)$.

5.4 Fórmula de Taylor II

Estableceremos en este apartado el desarrollo de Taylor de orden k cualquiera, comenzando, como antes, con un repaso de los resultados para funciones de una variable.

Sea $y = f(x)$ una función de una variable de clase C^k , $k \geq 2$, en un intervalo $I \subset \mathbb{R}$. Repitiendo el proceso que nos condujo a la fórmula (5.22), es fácil ver que el único C para el cual se tiene

$$f(x_0 + h) - f(x_0) = f'(x_0)h + \frac{1}{2}f''(x_0)h^2 + Ch^3 + o(h^3)$$

es $C = f'''(x_0)/6$, con lo que

$$f(x_0 + h) - f(x_0) = f'(x_0)h + \frac{1}{2}f''(x_0)h^2 + \frac{1}{6}f'''(x_0)h^3 + o(h^3) \quad (5.54)$$

Procediendo por inducción se obtiene la **fórmula de Taylor de orden k** con resto en forma asintótica:

$$f(x_0 + h) - f(x_0) = f'(x_0)h + \frac{1}{2}f''(x_0)h^2 + \dots + \frac{1}{k!}f^{(k)}(x_0)h^k + o(h^k) \quad (5.55)$$

Introduciendo la **diferencial p -ésima** de f en x_0 por

$$d^p f(x_0; h) \stackrel{\text{def}}{=} f^{(p)}(x_0)h^p$$

se tiene la **versión diferencial** de la fórmula de Taylor:

$$f(x_0 + h) - f(x_0) = df(x_0; h) + \frac{1}{2}d^2 f(x_0; h) + \dots + \frac{1}{k!}d^k f(x_0; h) + o(h^k) \quad (5.56)$$

El polinomio en h

$$P_k(x_0; h) = f(x_0) + f'(x_0)h + \frac{1}{2}f''(x_0)h^2 + \dots + \frac{1}{k!}f^{(k)}(x_0)h^k \quad (5.57)$$

es el **polinomio de Taylor de grado k** de f en x_0 . Se verifica

$$P_k(x_0; 0) = f(x_0), \quad P'_k(x_0; 0) = f'(x_0), \quad \dots, \quad P^{(k)}_k(x_0; 0) = f^{(k)}(x_0) \quad (5.58)$$

La cantidad

$$R_k(h) \equiv R_k(x_0; h) \stackrel{\text{def}}{=} f(x_0 + h) - P_k(x_0; h) \quad (5.59)$$

es el resto o término complementario de orden k (o, también, el error que se comete al sustituir $f(x_0 + h)$ por $P_k(x_0; h)$ en las proximidades de x_0). De él se conoce por la deducción anterior su comportamiento asintótico cuando $h \rightarrow 0$;

$$\lim_{h \rightarrow 0} \frac{R_k(h)}{h^k} = 0 \quad (5.60)$$

Descrito únicamente por esta propiedad, se dice que el resto está en **forma asintótica** o de **Peano**. Si se supone $f \in C^{k+1}(I)$ se puede obtener una expresión más explícita para $R_k(h)$. Veamos: se tiene, en virtud de (5.58) y (5.59),

$$R_k(0) = R'_k(0) = \dots = R_k^{(k)}(0) = 0$$

Aplicando el teorema de Cauchy a las funciones

$$R_k(t) = f(x_0 + t) - P_k(x_0; t) \text{ y } \varphi(t) = t^{k+1}$$

en el intervalo $t \in [0, h]$ se tendrá

$$\begin{aligned} \frac{R_k(h)}{\varphi(h)} &= \frac{R_k(h) - R_k(0)}{\varphi(h) - \varphi(0)} = \frac{R'_k(h_1)}{\varphi'(h_1)} = \frac{R'_k(h_1) - R'_k(0)}{\varphi'(h_1) - \varphi'(0)} = \\ &= \frac{R''_k(h_2)}{\varphi''(h_2)} = \dots = \frac{R_k^{(k)}(h_k)}{\varphi^{(k)}(h_k)} = \frac{R_k^{(k)}(h_k) - R_k^{(k)}(0)}{\varphi^{(k)}(h_k) - \varphi^{(k)}(0)} = \\ &= \frac{R_k^{(k+1)}(h_{k+1})}{\varphi^{(k+1)}(h_{k+1})} \end{aligned}$$

con $h_1 \in (0, h)$, $h_{j+1} \in (0, h_j)$, $j = 1, 2, \dots, k$. Pero $\varphi^{(k+1)}(t) = (k+1)!$, $R_k^{(k+1)}(t) = f^{(k+1)}(x_0 + t) - 0 = f^{(k+1)}(x_0 + t)$, de donde, poniendo $h_{k+1} = \theta h$ con $\theta \in (0, 1)$ (y así el resultado vale para h negativo), se deduce

$$R_k(h) = \frac{1}{(k+1)!} f^{(k+1)}(x_0 + \theta h) h^{k+1} \quad (5.61)$$

y en consecuencia

$$\begin{aligned} f(x_0 + h) - f(x_0) &= f'(x_0)h + \frac{1}{2}f''(x_0)h^2 + \dots + \frac{1}{k!}f^{(k)}(x_0)h^k + \\ &+ \frac{1}{(k+1)!}f^{(k+1)}(x_0 + \theta h)h^{k+1} \end{aligned} \quad (5.62)$$

con $0 < \theta < 1$. Esta es la fórmula de Taylor de orden k con **resto de Lagrange**. Hay otras expresiones "cuasi-explicitas" para el resto —que veremos

en la Nota 2 de final de capítulo— que, en ocasiones, son más útiles que ésta; pero la forma de Lagrange es quizá la más simple y, desde luego, la más fácil de recordar pues su estructura es la misma que la de los demás términos del desarrollo, con la derivada evaluada en un punto intermedio entre x_0 y $x_0 + h$, en lugar de en x_0 .

Para establecer la fórmula de Taylor general de funciones de n variables conviene disponer, antes de nada, de una notación cómoda para las derivadas parciales de cualquier orden de tales funciones:

Multíndices

Vefamos en el teorema 5.3 que si f es de clase C^p y las listas (i_1, i_2, \dots, i_p) y (j_1, j_2, \dots, j_p) contienen el mismo número α_1 de unos, el mismo número α_2 de doses, etc., entonces

$$\frac{\partial^p f}{\partial x_{i_1} \partial x_{i_2} \cdots \partial x_{i_p}} = \frac{\partial^p f}{\partial x_{j_1} \partial x_{j_2} \cdots \partial x_{j_p}} = \frac{\partial^p f}{\partial x_1^{\alpha_1} \partial x_2^{\alpha_2} \cdots \partial x_n^{\alpha_n}}$$

Pues bien, en relación a esta última notación para las derivadas parciales de orden p de una función f de clase C^p , ponemos:

Definición 5.2 Un **multíndice** de n componentes es una lista de n números naturales (incluyendo el cero)

$$\alpha = (\alpha_1, \alpha_2, \dots, \alpha_n)$$

El **peso** u **orden** de α es

$$|\alpha| \stackrel{\text{def}}{=} \alpha_1 + \alpha_2 + \cdots + \alpha_n$$

Sea $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_n)$ un **multíndice**. Entonces:

i) $\alpha! \stackrel{\text{def}}{=} \alpha_1! \alpha_2! \cdots \alpha_n!$

ii) Si $x = (x_1, x_2, \dots, x_n)$, se define

$$x^\alpha \stackrel{\text{def}}{=} x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_n^{\alpha_n}$$

iii) Si $f(x_1, x_2, \dots, x_n)$ es una función de n variables, la **derivada parcial de f de índice α** es

$$D^\alpha f \stackrel{\text{def}}{=} D_{x_1}^{\alpha_1} D_{x_2}^{\alpha_2} \cdots D_{x_n}^{\alpha_n} f = \frac{\partial^{|\alpha|} f}{\partial x_1^{\alpha_1} \partial x_2^{\alpha_2} \cdots \partial x_n^{\alpha_n}}$$

También se denota $D_\alpha f, \partial^\alpha f, \partial_\alpha f$, etc.

NOTAS. 1. Obsérvese la diferencia entre $\frac{\partial^{|\alpha|} f}{\partial x_1^{\alpha_1} \partial x_2^{\alpha_2} \dots \partial x_n^{\alpha_n}}$ y $\frac{\partial^p f}{\partial x_{i_1} \partial x_{i_2} \dots \partial x_{i_p}}$; en el primer caso, α_k representa el número de derivaciones respecto de la k -ésima variable x_k , mientras que, en el segundo, i_k indica cuál es la variable respecto de la que se efectúa la k -ésima derivada. Por ejemplo,

$$D^{(1,3,0,0)} f(x_1, x_2, x_3, x_4) = \frac{\partial^6}{\partial x_1 \partial x_2^3 \partial x_4} f(x_1, x_2, x_3, x_4)$$

que se escribiría

$$D_{1,3,2,4,4}^6 f \quad \text{ó} \quad D_{x_1 x_2 x_2 x_2 x_4 x_4}^6 f \quad \text{ó} \quad D_{x_1 x_2^3 x_4}^6 f$$

con la otra notación, supuesto que ése fuese el orden en que se toman las derivadas.

2. Si $p = |\alpha|$, el número $\frac{p!}{\alpha!}$ representa el número de listas ordenadas de p números del 1 al n en las cuales aparecen α_1 unos, α_2 doses, ..., α_n "enes". En general, es el número de permutaciones distintas que se pueden formar con p objetos entre los cuales hay α_1 iguales entre sí, otros α_2 iguales entre sí, etc.

Para apreciar mejor la diferencia entre ambas notaciones, así como sus ventajas e inconvenientes respectivos, expresemos simbólicamente la potencia p -ésima de una suma de n términos en la primera forma:

$$(a_1 + a_2 + \dots + a_n)^p = \sum_{i_1, i_2, \dots, i_p=1}^n a_{i_1} a_{i_2} \dots a_{i_p}$$

donde la suma se extiende a todas las listas (i_1, i_2, \dots, i_p) de p números del 1 al n . Por la propiedad conmutativa, coincidirán todos los términos en los cuales haya el mismo número α_1 de unos, α_2 de doses, etc. El número de dichas listas es, como hemos indicado antes, $p!/\alpha!$. Por tanto,

$$\begin{aligned} (a_1 + a_2 + \dots + a_n)^p &= \sum_{\alpha_1 + \alpha_2 + \dots + \alpha_n = p} \frac{p!}{\alpha_1! \alpha_2! \dots \alpha_n!} a_1^{\alpha_1} a_2^{\alpha_2} \dots a_n^{\alpha_n} = \\ &= \sum_{|\alpha|=p} \frac{p!}{\alpha!} a^\alpha \end{aligned}$$

Esta simplificación que proporciona la notación con multiíndices será muy útil para escribir de una manera compacta expresiones en las que intervienen derivadas parciales de cualquier orden, como vamos a apreciar a continuación en la fórmula de Taylor en varias variables.

Para obtener dicha fórmula, utilizaremos de nuevo la función auxiliar

$$F(t) = f(x_0 + th)$$

de una variable y aplicaremos a ella la fórmula de Taylor que acabamos de deducir. Necesitamos para ello calcular sus derivadas sucesivas:

Teorema 5.5 Sean $D \subset \mathbb{R}^n$ abierto, $x_0 \in D$ y $h \in \mathbb{R}^n$ tales que el segmento $[x_0, x_0 + h]$ está contenido en D y sea $f \in C^k(D)$. Entonces, la función

$$F(t) \stackrel{\text{def}}{=} f(x_0 + th)$$

es de clase C^k y se tiene

$$F^{(p)}(t) = \sum_{i_1, i_2, \dots, i_p=1}^n D_{x_{i_1} x_{i_2} \dots x_{i_p}}^p f(x_0 + th) h_{i_1} h_{i_2} \dots h_{i_p} \quad (5.63)$$

para todo $p \leq k$, o, en la notación de multiíndices,

$$\begin{aligned} F^{(p)}(t) &= \\ &= \sum_{\alpha_1 + \alpha_2 + \dots + \alpha_n = p} \frac{p!}{\alpha_1! \alpha_2! \dots \alpha_n!} \frac{\partial^p f}{\partial x_1^{\alpha_1} \partial x_2^{\alpha_2} \dots \partial x_n^{\alpha_n}}(x_0 + th) h_1^{\alpha_1} h_2^{\alpha_2} \dots h_n^{\alpha_n} = \\ &= \sum_{|\alpha|=p} \frac{p!}{\alpha!} D^\alpha f(x_0 + th) h^\alpha \end{aligned} \quad (5.64)$$

NOTA. Introduciendo el operador simbólico

$$d(f) \stackrel{\text{def}}{=} (h_1 D_{x_1} + h_2 D_{x_2} + \dots + h_n D_{x_n})(f)$$

podemos también escribir:

$$F^{(p)}(t) = (h_1 D_{x_1} + h_2 D_{x_2} + \dots + h_n D_{x_n})^p (f)(x_0 + th) = d^p(f)(x_0 + th) \quad (5.65)$$

donde la potencia simbólica

$$(h_1 D_{x_1} + h_2 D_{x_2} + \dots + h_n D_{x_n})^p$$

se utiliza (por analogía con la fórmula multinomial vista en el apartado 1 de este capítulo) para representar el operador diferencial

$$\begin{aligned} &(h_1 D_{x_1} + h_2 D_{x_2} + \dots + h_n D_{x_n})^p = \\ &= \sum_{\alpha_1 + \alpha_2 + \dots + \alpha_n = p} \frac{p!}{\alpha_1! \alpha_2! \dots \alpha_n!} (h_1 D_{x_1})^{\alpha_1} (h_2 D_{x_2})^{\alpha_2} \dots (h_n D_{x_n})^{\alpha_n} = \\ &= \sum_{|\alpha|=p} \frac{p!}{\alpha!} h^\alpha D^\alpha \end{aligned}$$

Por ejemplo, si $n = 2$

$$(hD_x + kD_y)^p f = \sum_{r+s=p} \frac{p!}{r!s!} h^r k^s D_x^r D_y^s f = \sum_{r=0}^p \binom{p}{r} h^r k^{p-r} D_x^r D_y^{p-r} f$$

y así:

$$\begin{aligned} (hD_x + kD_y)^2 f &= h^2 \frac{\partial^2 f}{\partial x^2} + 2hk \frac{\partial^2 f}{\partial x \partial y} + k^2 \frac{\partial^2 f}{\partial y^2} \\ (hD_x + kD_y)^3 f &= h^3 \frac{\partial^3 f}{\partial x^3} + 3h^2 k \frac{\partial^3 f}{\partial x^2 \partial y} + 3hk^2 \frac{\partial^3 f}{\partial x \partial y^2} + k^3 \frac{\partial^3 f}{\partial y^3} \end{aligned}$$

Demostración. Procedemos por inducción. Para $p = 1$, se trata de la fórmula de derivación de funciones compuestas:

$$F'(t) = \sum_{i=1}^n D_{x_i} f(x_0 + th) h_i = \left(\sum_{i=1}^n h_i D_{x_i} \right) f(x_0 + th) \quad (5.66)$$

(ya lo hemos visto también para $p = 2$). Veamos que de la validez de (5.64) para $p - 1$ se deduce la validez para p . Por la hipótesis de inducción:

$$F^{(p-1)}(t) = \sum_{|\beta|=p-1} \frac{(p-1)!}{\beta!} D^\beta f(x_0 + th) h^\beta$$

Aplicando (5.66) a cada una de las derivadas $D^\beta f(x_0 + th)$, $|\beta| = p - 1$, se tiene

$$\begin{aligned} \frac{d}{dt} \left[D^\beta f(x_0 + th) h^\beta \right] &= h^\beta \frac{d}{dt} \left[D^\beta f(x_0 + th) \right] = \\ &= h^\beta \sum_{i=1}^n D_{x_i} D^\beta f(x_0 + th) h_i = \\ &= \sum_{i=1}^n h_1^{\beta_1} \dots h_i^{\beta_i+1} \dots h_n^{\beta_n} D_{x_1}^{\beta_1} \dots D_{x_i}^{\beta_i+1} \dots D_{x_n}^{\beta_n} f(x_0 + th) \end{aligned}$$

y entonces

$$\begin{aligned} F^{(p)}(t) &= \\ &= \sum_{|\beta|=p-1} \frac{(p-1)!}{\beta!} \left(\sum_{i=1}^n h_1^{\beta_1} \dots h_i^{\beta_i+1} \dots h_n^{\beta_n} D_{x_1}^{\beta_1} \dots D_{x_i}^{\beta_i+1} \dots D_{x_n}^{\beta_n} f(x_0 + th) \right) = \\ &= \sum_{|\alpha|=p} a_\alpha D^\alpha f(x_0 + th) h^\alpha \end{aligned}$$

con

$$\begin{aligned} a_\alpha &= \sum_{i=1}^n \frac{(p-1)!}{\alpha_1! \dots (\alpha_i - 1)! \dots \alpha_n!} = \\ &= \frac{(p-1)!}{(\alpha_1 - 1)! \alpha_2! \dots \alpha_n!} + \frac{(p-1)!}{\alpha_1! (\alpha_2 - 1)! \dots \alpha_n!} + \dots + \frac{(p-1)!}{\alpha_1! \dots \alpha_{n-1}! (\alpha_n - 1)!} = \\ &= \frac{\alpha_1 (p-1)!}{\alpha!} + \frac{\alpha_2 (p-1)!}{\alpha!} + \dots + \frac{\alpha_n (p-1)!}{\alpha!} = \frac{p!}{\alpha!} \end{aligned}$$

según queríamos demostrar.

Las diferenciales sucesivas de f se obtienen particularizando estas fórmulas para $t = 0$, es decir, aplicando p veces el operador de diferenciación d (de donde la notación d^p):

Definición 5.3 Se llama *diferencial de orden p de f* a la expresión

$$\begin{aligned} d^p f(x_0; h) &\stackrel{\text{def}}{=} \left(\sum_{i=1}^n h_i D_{x_i} \right)^p (f)(x_0) = \\ &= \sum_{i_1, i_2, \dots, i_p=1}^n D_{x_{i_1} x_{i_2} \dots x_{i_p}} f(x_0) h_{i_1} h_{i_2} \dots h_{i_p} = \\ &= \sum_{|\alpha|=p} \frac{p!}{\alpha!} D_{\alpha} f(x_0) h^{\alpha} \end{aligned}$$

Obsérvese que $d^p f(x_0; h)$, como función de h para x_0 fijo, es un polinomio homogéneo de grado p .

Teorema 5.6 (Fórmula de Taylor para una función de n variables)

Sean $D \subset \mathbb{R}^n$ abierto, $x_0 \in D$ y $h \in \mathbb{R}^n$ tales que el segmento $[x_0, x_0 + h]$ está contenido en D .

1. Si $f \in C^k(D)$, entonces:

$$f(x_0 + h) = f(x_0) + df(x_0; h) + \frac{1}{2!} d^2 f(x_0; h) + \dots + \frac{1}{k!} d^k f(x_0; h) + R_k(h) \quad (5.67)$$

donde $R_k(h) = o(|h|^k)$ (forma asintótica del resto).

2. Si $f \in C^{k+1}(D)$, se tiene (5.67) con

$$R_k(h) = \frac{1}{(k+1)!} d^{k+1} f(x_0 + \theta h; h) \quad (\text{forma de Lagrange})$$

siendo θ un número entre 0 y 1.

En forma desarrollada,

$$\begin{aligned} f(x_0 + h) &= f(x_0) + \left(\sum_{i=1}^n h_i D_{x_i} \right) f(x_0) + \frac{1}{2!} \left(\sum_{i=1}^n h_i D_{x_i} \right)^2 f(x_0) + \dots + \\ &\quad + \frac{1}{k!} \left(\sum_{i=1}^n h_i D_{x_i} \right)^k f(x_0) + R_k(h) \end{aligned}$$

o bien

$$\begin{aligned} f(x_0 + h) &= f(x_0) + \sum_{i=1}^n D_{x_i} f(x_0) h_i + \frac{1}{2!} \sum_{i,j=1}^n D_{x_i x_j}^2 f(x_0) h_i h_j + \dots + \\ &\quad + \frac{1}{k!} \sum_{i_1, i_2, \dots, i_k=1}^n D_{x_{i_1} x_{i_2} \dots x_{i_k}}^k f(x_0) h_{i_1} h_{i_2} \dots h_{i_k} + R_k(h) \end{aligned}$$

donde, si $f \in C^{k+1}(D)$, se tiene, para cierto $\theta \in (0, 1)$,

$$\begin{aligned} R_k &= \frac{1}{(k+1)!} \left(\sum_{i=1}^n h_i D_{x_i} \right)^{k+1} f(x_0 + \theta h) = \\ &= \frac{1}{(k+1)!} \left(\sum_{i_1, i_2, \dots, i_{k+1}=1}^n D_{x_{i_1} x_{i_2} \dots x_{i_{k+1}}}^{k+1} f(x_0 + \theta h) h_{i_1} h_{i_2} \dots h_{i_{k+1}} \right) \end{aligned}$$

En la notación de multíndices,

$$f(x_0 + h) = f(x_0) + \sum_{|\alpha| \leq k} \frac{1}{\alpha!} D_\alpha f(x_0) h^\alpha + R_k(h) \quad (5.68)$$

donde, para cierto $\theta \in (0, 1)$,

$$R_k(h) = \sum_{|\alpha|=k+1} \frac{1}{\alpha!} D^\alpha f(x_0 + \theta h) h^\alpha$$

El polinomio de Taylor de grado k de f en x_0 es

$$\begin{aligned} P_k(h) &\stackrel{\text{def}}{=} f(x_0) + \sum_{i=1}^n D_{x_i} f(x_0) h_i + \frac{1}{2!} \sum_{i,j=1}^n D_{x_i x_j}^2 f(x_0) h_i h_j + \dots + \\ &+ \frac{1}{k!} \sum_{i_1, i_2, \dots, i_k=1}^n D_{x_{i_1} x_{i_2} \dots x_{i_k}}^k f(x_0) h_{i_1} h_{i_2} \dots h_{i_k} = \\ &= f(x_0) + \sum_{|\alpha| \leq k} \frac{1}{\alpha!} D^\alpha f(x_0) h^\alpha \end{aligned} \quad (5.69)$$

La idea de la demostración es la misma que para el caso $k=2$: se parte del desarrollo de Taylor de orden $k-1$ de la función $F(t)$ con resto de Lagrange:

$$F(1) = F(0) + F'(0) + \frac{1}{2} F''(0) + \dots + \frac{1}{(k-1)!} F^{(k-1)}(0) + \frac{1}{k!} F^{(k)}(\theta),$$

$0 < \theta < 1$, para obtener, utilizando el teorema 5.5,

$$\begin{aligned} f(x_0 + h) &= f(x_0) + df(x_0; h) + \frac{1}{2!} d^2 f(x_0; h) + \dots + \\ &+ \frac{1}{(k-1)!} d^{(k-1)} f(x_0; h) + \frac{1}{k!} d^k f(x_0 + \theta h; h) \end{aligned}$$

(Desarrollo de Taylor de $f(x)$ de orden $k-1$ con resto de Lagrange). Se tiene

$$\frac{f(x_0+h) - P_k(h)}{|h|^k} = \frac{1}{k!|h|^k} \left[d^k f(x_0 + \theta h; h) - d^k f(x_0; h) \right]$$

Como las derivadas de orden k son continuas, por hipótesis, podemos escribir para cada una de ellas, $D^\alpha f$, $|\alpha| = k$,

$$D^\alpha f(x_0 + \theta h) = D^\alpha f(x_0) + \delta_\alpha(h)$$

con $\delta_\alpha(h) \rightarrow 0$ cuando $|h| \rightarrow 0$. Entonces, dado que

$$d^k f(x_0 + \theta h; h) = \sum_{|\alpha|=k} \frac{k!}{\alpha!} D^\alpha f(x_0 + \theta h) h^\alpha$$

tenemos:

$$d^k f(x_0 + \theta h; h) = d^k f(x_0; h) + \sum_{|\alpha|=k} \frac{k!}{\alpha!} \delta_\alpha(h) h^\alpha$$

El segundo término es $o(|h|^k)$, pues cada monomio h^α se puede acotar superiormente por $|h|^k$, y los factores $\delta_\alpha(h)$ tienden a 0:

$$\begin{aligned} |h^\alpha| &= |h_1^{\alpha_1} h_2^{\alpha_2} \dots h_n^{\alpha_n}| = |h_1|^{\alpha_1} |h_2|^{\alpha_2} \dots |h_n|^{\alpha_n} \leq \\ &\leq |h|^{\alpha_1 + \alpha_2 + \dots + \alpha_n} = |h|^k \end{aligned}$$

y entonces

$$\begin{aligned} \frac{1}{k!} \frac{|d^k f(x_0 + \theta h; h) - d^k f(x_0; h)|}{|h|^k} &= \frac{1}{k!} \frac{\left| \sum_{|\alpha|=k} \frac{k!}{\alpha!} \delta_\alpha(h) h^\alpha \right|}{|h|^k} \leq \\ &\leq \sum_{|\alpha|=k} \frac{1}{\alpha!} |\delta_\alpha(h)| \xrightarrow{h \rightarrow 0} 0 \end{aligned}$$

de donde

$$\frac{1}{|h|^p} \left| \sum_{|\alpha|=p} \frac{p!}{\alpha!} \delta_\alpha(h) h^\alpha \right| \leq \sum_{|\alpha|=p} \frac{p!}{\alpha!} |\delta_\alpha(h)| \xrightarrow{h \rightarrow 0} 0$$

La demostración del punto 2 del teorema deriva directamente del desarrollo de Taylor de $F(t)$ de orden k con resto de Lagrange.

Escribamos explícitamente para el caso $n = 2$ el comienzo del desarrollo de Taylor:

$$\begin{aligned} f(x_0 + h, y_0 + k) &= f(x_0, y_0) + [hf'_x(x_0, y_0) + kf'_y(x_0, y_0)] + \\ &+ \frac{1}{2} [h^2 f''_{xx}(x_0, y_0) + 2hkf''_{xy}(x_0, y_0) + k^2 f''_{yy}(x_0, y_0)] + \\ &+ \frac{1}{6} [h^3 f'''_{xxx}(x_0, y_0) + 3h^2 k f'''_{xxy}(x_0, y_0) + 3hk^2 f'''_{xyy}(x_0, y_0) + k^3 f'''_{yyy}(x_0, y_0)] + \dots \end{aligned}$$

Poniendo $x = x_0 + h$, el polinomio de Taylor de grado k , en la notación de multiíndices, es

$$P_k(x) = f(x_0) + \sum_{|\alpha| \leq k} \frac{1}{\alpha!} D^\alpha f(x_0) (x - x_0)^\alpha$$

y las fórmulas de Taylor de orden k :

$$\begin{aligned} f(x) &= f(x_0) + \sum_{|\alpha| \leq k} \frac{1}{\alpha!} D^\alpha f(x_0) (x - x_0)^\alpha + o(|x - x_0|^k), \\ &\quad (\text{con resto de Peano}) \end{aligned}$$

y

$$\begin{aligned} f(x) &= f(x_0) + \sum_{|\alpha| \leq k} \frac{1}{\alpha!} D^\alpha f(x_0) (x - x_0)^\alpha + \sum_{|\alpha| = k+1} \frac{1}{\alpha!} D^\alpha f(X) (x - x_0)^\alpha \\ X &= x_0 + \theta(x - x_0), \text{ para un } \theta \in (0, 1) \text{ (con resto de Lagrange)} \end{aligned}$$

La principal aplicación de la expresión de Lagrange para el término complementario es disponer de una estimación del error que se comete al aproximar $f(x)$ por su polinomio de Taylor $P_k(x)$:

Teorema 5.7 Sean $D \subset \mathbb{R}^n$ abierto, $x_0 \in D$ y $h \in \mathbb{R}^n$ tales que el segmento $[x_0, x_0 + h]$ está contenido en D y sean $f \in C^{k+1}(D)$ y $P_k(x)$ su polinomio de Taylor de grado k en x_0 . Entonces, si M es una cota superior de todas las derivadas parciales de orden $k+1$ de f en el segmento $[x_0, x_0 + h]$ se tiene:

$$|f(x) - P_k(x)| \leq \frac{M}{(k+1)!} (|x_1 - x_1^0| + \dots + |x_n - x_n^0|)^{k+1}$$

para todo $x \in [x_0, x_0 + h]$.

En efecto, se tiene, con $h = x - x_0$

$$\begin{aligned} |R_k(h)| &= \left| \sum_{|\alpha|=k+1} \frac{1}{\alpha!} D^\alpha f(x_0 + \theta h) h^\alpha \right| \leq \\ &\leq \frac{M}{(k+1)!} \sum_{|\alpha|=k+1} \frac{(p+1)!}{\alpha!} |h_1|^{\alpha_1} |h_2|^{\alpha_2} \dots |h_n|^{\alpha_n} = \\ &= \frac{M}{(k+1)!} (|h| + |h_2| + \dots + |h_n|)^{k+1} \end{aligned}$$

Unicidad del polinomio de Taylor

La propiedad de diferenciabilidad de una función f en el punto x_0 implica que la única aplicación lineal

$$\mathbb{R}^n \ni h \mapsto a_1 h_1 + \dots + a_n h_n \in \mathbb{R}, \quad a_i \in \mathbb{R}$$

tal que

$$\frac{f(x_0 + h) - f(x_0) - [a_1 h_1 + \dots + a_n h_n]}{|h|} \xrightarrow{h \rightarrow 0} 0$$

es la que corresponde a $a_i = f'_{x_i}(x_0)$. De manera análoga, el único polinomio de grado k que aproxima a $f(x_0 + h)$ con un error $o(|h|^k)$ es el polinomio de Taylor de orden k . Veamos:

Lema 5.1 Sean $Q(h)$ y $Q^*(h)$ dos polinomios en $h = (h_1, \dots, h_n)$ de grado k tales que

$$\lim_{h \rightarrow 0} \frac{Q(h) - Q^*(h)}{|h|^k} = 0 \quad (5.70)$$

Entonces $Q = Q^*$.

Demostración. Supongamos que, por el contrario, se tiene $Q \neq Q^*$. Sea $Q(h) - Q^*(h) = F(h) + G(h)$, donde $F(h)$ es el polinomio homogéneo formado por todos los términos del menor grado q que aparece en $Q(h) - Q^*(h)$ y $G(h)$ el polinomio formado por todos los términos de grado mayor que q :

$$F(h) = \sum_{|\alpha|=q} a_\alpha h^\alpha, \quad G(h) = \sum_{|\alpha|>q} b_\alpha h^\alpha$$

Tomemos $\bar{h} \neq 0$ tal que $F(\bar{h}) \neq 0$, lo que posible ya que $F(h) \neq 0$. Si fuese $q = k$, se tendría $G(h) \equiv 0$ y, para $t > 0$,

$$\lim_{t \rightarrow 0} \frac{Q(t\bar{h}) - Q^*(t\bar{h})}{|t\bar{h}|^k} = \lim_{t \rightarrow 0} \frac{t^k F(\bar{h})}{t^k |\bar{h}|^k} = \frac{F(\bar{h})}{|\bar{h}|^k} \neq 0$$

lo que contradice (5.70). Si $q < k$ se tiene, para t suficientemente pequeño,

$$|t\bar{h}|^q = t^q |t\bar{h}|^q \geq t^k |\bar{h}|^k = |t\bar{h}|^k$$

y, por (5.70), ha de verificarse

$$\lim_{t \rightarrow 0} \frac{Q(t\bar{h}) - Q^*(t\bar{h})}{|t\bar{h}|^q} = 0$$

Pero

$$\lim_{t \rightarrow 0} \frac{Q(t\bar{h}) - Q^*(t\bar{h})}{|t\bar{h}|^q} = \lim_{t \rightarrow 0} \frac{F(t\bar{h}) + G(t\bar{h})}{t^q |\bar{h}|^q} =$$

$$\lim_{t \rightarrow 0} \frac{t^q F(\bar{h})}{t^q |\bar{h}|^q} + \lim_{t \rightarrow 0} \frac{G(t\bar{h})}{t^q |\bar{h}|^q} = \frac{F(\bar{h})}{|\bar{h}|^q} \neq 0$$

ya que todos los términos de G son de grado mayor que q . Se tiene, pues, una contradicción con la hipótesis (5.70) y ha de ser $Q = Q^*$.

A partir de este lema se deduce sin dificultad la unicidad del polinomio de Taylor $P_k(h)$. Pues si $Q(h)$ es otro polinomio de grado k tal que

$$\lim_{h \rightarrow 0} \frac{f(x_0 + h) - Q(h)}{|h|^k} = 0 \quad (5.71)$$

entonces

$$\left| \frac{Q(h) - P_k(h)}{|h|^k} \right| \leq \left| \frac{f(x_0 + h) - Q(h)}{|h|^k} \right| + \left| \frac{f(x_0 + h) - P_k(h)}{|h|^k} \right| \xrightarrow{h \rightarrow 0} 0$$

por (5.71), de donde, por el lema, $Q = P_k$.

Este resultado de unicidad permite determinar polinomios de Taylor sin necesidad de derivar, construyéndolos a partir de otros más simples ya conocidos y comprobando únicamente que satisfacen la propiedad (5.71):

EJEMPLO. Determinemos el polinomio de Taylor de grado 4 en $(0,0)$ de la función $f(x, y) = e^x \cos y$. Se tiene

$$\begin{aligned} e^x &= 1 + x + \frac{x^2}{2} + \frac{x^3}{6} + \frac{x^4}{24} + R(x), \text{ con } \lim_{x \rightarrow 0} \frac{R(x)}{x^4} = 0 \\ \cos y &= 1 - \frac{y^2}{2} + \frac{y^4}{24} + R(y), \text{ con } \lim_{y \rightarrow 0} \frac{R(y)}{y^4} = 0 \end{aligned}$$

Multiplicando ambos polinomios tenemos

$$e^x \cos y = 1 + x + \frac{1}{2}x^2 - \frac{1}{2}y^2 + \frac{1}{6}x^3 - \frac{1}{2}xy^2 + \\ + \frac{1}{24}x^4 - \frac{1}{4}x^2y^2 + \frac{1}{24}y^4 + R(x, y)$$

donde

$$R(x, y) = \frac{1}{24}xy^4 + \frac{1}{48}x^2y^4 + \frac{1}{6}x^3\left(-\frac{y^2}{2} + \frac{y^4}{24}\right) + \frac{1}{24}x^4\left(-\frac{y^2}{2} + \frac{y^4}{24}\right) + \\ + \left(1 + x + \frac{x^2}{2} + \frac{x^3}{6} + \frac{x^4}{24}\right)R(y) + \left(1 - \frac{y^2}{2} + \frac{y^4}{24}\right)R(x) + R(x)R(y)$$

Puesto que (compruébese como ejercicio)

$$\lim_{(x,y) \rightarrow (0,0)} \frac{R(x,y)}{(x^2+y^2)^2} = 0$$

se tiene que

$$1 + x + \frac{1}{2}x^2 - \frac{1}{2}y^2 + \frac{1}{6}x^3 - \frac{1}{2}xy^2 + \frac{1}{24}x^4 - \frac{1}{4}x^2y^2 + \frac{1}{24}y^4$$

es el polinomio de Taylor de grado 4 de $f(x, y) = e^x \cos y$ en $(0, 0)$.

5.5 Notas y complementos

1. Cambio del orden de derivación

El problema del cambio del orden de derivación es, en realidad, una cuestión de *límites iterados*. En efecto, observemos primero que

$$f''_{xy}(x_0, y_0) = \lim_{k \rightarrow 0} \frac{f'_x(x_0, y_0 + k) - f'_x(x_0, y_0)}{k} = \\ = \lim_{k \rightarrow 0} \frac{1}{k} \left(\lim_{h \rightarrow 0} \frac{f(x_0 + h, y_0 + k) - f(x_0, y_0 + k)}{h} - \right. \\ \left. - \lim_{h \rightarrow 0} \frac{f(x_0 + h, y_0) - f(x_0, y_0)}{h} \right) = \\ = \lim_{k \rightarrow 0} \lim_{h \rightarrow 0} \frac{1}{hk} [f(x_0 + h, y_0 + k) - f(x_0, y_0 + k) - f(x_0 + h, y_0) + f(x_0, y_0)]$$

y, análogamente,

$$f''_{yx}(x_0, y_0) = \\ = \lim_{h \rightarrow 0} \lim_{k \rightarrow 0} \frac{1}{hk} [f(x_0 + h, y_0 + k) - f(x_0 + h, y_0) - f(x_0, y_0 + k) + f(x_0, y_0)]$$

Obsérvese que las expresiones entre corchetes coinciden en ambos casos. Llámoslas

$$A(h, k) \stackrel{\text{def}}{=} f(x_0 + h, y_0 + k) - f(x_0 + h, y_0) - f(x_0, y_0 + k) + f(x_0, y_0)$$

El problema se reduce así a decidir bajo qué hipótesis se puede garantizar la igualdad de los límites iterados:

$$\lim_{k \rightarrow 0} \lim_{h \rightarrow 0} \frac{1}{hk} A(h, k) = \lim_{h \rightarrow 0} \lim_{k \rightarrow 0} \frac{1}{hk} A(h, k)$$

siempre que los "límites parciales" $\lim_{h \rightarrow 0} [A(h, k)/hk]$ y $\lim_{k \rightarrow 0} [A(h, k)/hk]$ en la región $R = \{(h, k) : h \neq 0, k \neq 0\}$ estén bien definidos, como sucede aquí. Se puede probar que para la igualdad de límites iterados es condición suficiente la existencia del límite doble $\lim_{(h,k) \rightarrow (0,0)} [A(h, k)/hk]$ en R . Este límite doble existe bajo ciertas hipótesis, la más importante de las cuales es la que hemos utilizado en el apartado 1:

Lema 5.2 Si f''_{xy} existe en un entorno de (x_0, y_0) y es continua en (x_0, y_0) , entonces

$$\begin{aligned} f''_{xy}(x_0, y_0) &= \\ &= \lim_{(h,k) \rightarrow (0,0)} \frac{f(x_0 + h, y_0 + k) - f(x_0 + h, y_0) - f(x_0, y_0 + k) + f(x_0, y_0)}{hk} \end{aligned}$$

Demostración. Sean (h, k) lo suficientemente pequeños como para que el rectángulo R_{hk} de vértices opuestos (x_0, y_0) y $(x_0 + h, y_0 + k)$ esté contenido en el entorno de (x_0, y_0) en el cual existe la derivada f''_{xy} . Recordemos que

$$A(h, k) \stackrel{\text{def}}{=} f(x_0 + h, y_0 + k) - f(x_0 + h, y_0) - f(x_0, y_0 + k) + f(x_0, y_0)$$

y definamos

$$F(x) \stackrel{\text{def}}{=} f(x, y_0 + k) - f(x, y_0)$$

Entonces, $F'(x) = f'_x(x, y_0 + k) - f'_x(x, y_0)$ y $A(h, k)$ se puede escribir de las dos formas equivalentes:

$$\begin{aligned} A(h, k) &= F(x_0 + h) - F(x_0) = \\ &= hF'(x_0 + \theta_1 h) = h[f'_x(x_0 + \theta_1 h, y_0 + k) - f'_x(x_0 + \theta_1 h, y_0)] \end{aligned}$$

donde $0 < \theta_1 < 1$. Volviendo a aplicar dicho teorema,

$$f'_x(x_0 + \theta_1 h, y_0 + k) - f'_x(x_0 + \theta_1 h, y_0) = kf''_{xy}(x_0 + \theta_1 h, y_0 + \theta_2 k),$$

donde $0 < \theta_2 < 1$. Por tanto,

$$A(h, k) = hk f''_{xy}(x_0 + \theta_1 h, y_0 + \theta_2 k)$$

Por ser f''_{xy} continua en (x_0, y_0) , vemos que

$$f''_{xy}(x_0, y_0) = \lim_{(h,k) \rightarrow (0,0)} \frac{1}{hk} A(h, k)$$

Como consecuencia, tenemos la siguiente mejora notable del resultado (debido a BONNET) del apartado 1:

Teorema 5.8 (SCHWARZ): Si f'_x, f'_y y f''_{xy} existen en un entorno de (x_0, y_0) , y f''_{xy} es continua en (x_0, y_0) , entonces $f''_{yx}(x_0, y_0)$ existe y coincide con $f''_{xy}(x_0, y_0)$.

En este caso no hace falta exigir nada acerca de f''_{xy} (ni siquiera su existencia, que es parte de la conclusión del teorema).

Demostración. Por el lema anterior, sabemos que

$$f''_{xy}(x_0, y_0) = \lim_{(h,k) \rightarrow (0,0)} \frac{1}{hk} A(h, k)$$

Por otra parte, la existencia de f'_y garantiza la existencia de

$$\begin{aligned} \lim_{k \rightarrow 0} \frac{f(x_0 + h, y_0 + k) - f(x_0 + h, y_0) - f(x_0, y_0 + k) + f(x_0, y_0)}{hk} &= \\ &= \frac{f'_y(x_0 + h, y_0) - f'_y(x_0, y_0)}{h} \end{aligned}$$

Es decir, para cada $h \neq 0$ existe $B(h) \stackrel{\text{def}}{=} \lim_{k \rightarrow 0} \frac{1}{hk} A(h, k)$. Por tanto, la teoría general de límites garantiza la existencia de $\lim_{h \rightarrow 0} B(h)$, que debe además coincidir con el límite doble, es decir, con $f''_{xy}(x_0, y_0)$.

Recordemos brevemente la demostración de este último hecho: Supongamos que existe el límite doble l de $\frac{A(h, k)}{hk}$ cuando $(h, k) \rightarrow (0, 0)$ siendo $h, k \neq 0$. Esto significa que para todo $\varepsilon > 0$ existe $\delta > 0$ tal que $0 < |h|, |k| < \delta$ implica $\left| \frac{A(h, k)}{hk} - l \right| < \varepsilon$. Entonces, supuesto existente $\tilde{A}(h) = \lim_{k \rightarrow 0} \frac{A(h, k)}{k}$, tendríamos

$$\left| \frac{A(h, k) - lh}{k} \right| < \varepsilon |h| \Rightarrow \left| \tilde{A}(h) - lh \right| \leq \varepsilon |h|$$

de donde

$$|B(h) - l| = \left| \frac{\tilde{A}(h)}{h} - l \right| \leq \varepsilon$$

como queríamos demostrar.

Un resultado basado en hipótesis algo distintas es el siguiente

Teorema 5.9 (HEFFTER-YOUNG) Si f'_x y f'_y existen en un entorno de (x_0, y_0) y son diferenciables en (x_0, y_0) , entonces existen $f''_{xy}(x_0, y_0)$ y $f''_{yx}(x_0, y_0)$, y coinciden.

Demostración. Se basa también en demostrar la existencia del límite doble. Partimos de la relación ya conocida

$$A(h, k) = h [f'_x(x_0 + \theta_1 h, y_0 + k) - f'_x(x_0 + \theta_1 h, y_0)]$$

y usamos la diferenciability de f'_x en (x_0, y_0) :

$$f'_x(x_0 + h, y_0 + k) - f'_x(x_0, y_0) = Mh + Nk + \delta_1(m, n)$$

donde $M = f''_{xx}(x_0, y_0)$, $N = f''_{xy}(x_0, y_0)$ y $\delta_1(m, n) = o(|(m, n)|)$

En particular,

$$\begin{aligned} f'_x(x_0 + \theta_1 h, y_0 + k) - f'_x(x_0, y_0) &= M\theta_1 h + Nk + \delta_1(\theta_1 h, k) \\ f'_x(x_0 + \theta_1 h, y_0) - f'_x(x_0, y_0) &= M\theta_1 h + \delta_1(\theta_1 h, 0) \end{aligned}$$

Por tanto,

$$\begin{aligned} A(h, k) &= h [f'_x(x_0 + \theta_1 h, y_0 + k) - f'_x(x_0 + \theta_1 h, y_0)] = \\ &= h [f''_{xy}(x_0, y_0)k + \delta_1(\theta_1 h, k) - \delta_1(\theta_1 h, 0)] \end{aligned}$$

Análogamente,

$$f'_y(x_0 + m, y_0 + n) - f'_y(x_0, y_0) = Pm + Qn + \delta_2(m, n)$$

donde $P = f''_{yx}(x_0, y_0)$, $Q = f''_{yy}(x_0, y_0)$ y $\delta_2(m, n) = o(|(m, n)|)$. Igual que antes,

$$A(h, k) = k [f'_y(x_0 + h, y_0 + \theta_2 k) - f'_y(x_0, y_0 + \theta_2 k)]$$

y como

$$\begin{aligned} f'_y(x_0 + h, y_0 + \theta_2 k) - f'_y(x_0, y_0) &= Ph + Q\theta_2 k + \delta_2(h, \theta_2 k) \\ f'_y(x_0, y_0 + \theta_2 k) - f'_y(x_0, y_0) &= Q\theta_2 k + \delta_2(0, \theta_2 k) \end{aligned}$$

deducimos que

$$\begin{aligned} A(h, k) &= k [f'_y(x_0 + h, y_0 + \theta_2 k) - f'_y(x_0, y_0 + \theta_2 k)] = \\ &= k [f''_{yx}(x_0, y_0)h + \delta_2(h, \theta_2 k) - \delta_2(0, \theta_2 k)] \end{aligned}$$

Haciendo ahora $h = k$,

$$\begin{aligned} A(h, h) &= h [f''_{xy}(x_0, y_0)h + \delta_1(\theta_1 h, h) - \delta_1(\theta_1 h, 0)] \\ &\quad y \\ A(h, h) &= h [f''_{yx}(x_0, y_0)h + \delta_2(h, \theta_2 h) - \delta_2(0, \theta_2 h)] \end{aligned}$$

de donde

$$\begin{aligned} &f''_{xy}(x_0, y_0) - f''_{yx}(x_0, y_0) = \\ &= \frac{\delta_1(\theta_1 h, h) - \delta_1(\theta_1 h, 0) - (\delta_2(h, \theta_2 h) - \delta_2(0, \theta_2 h))}{h} \end{aligned}$$

y esta última cantidad tiende a 0 cuando $h \rightarrow 0$. Por tanto, $f''_{xy}(x_0, y_0) - f''_{yx}(x_0, y_0) = 0$ como queríamos demostrar.

Se puede comprobar que la función

$$f(x, y) = xy \frac{x^2 - y^2}{x^2 + y^2} \quad \text{si } (x, y) \neq (0, 0), \quad f(0, 0) = 0$$

del ejemplo dado en el apartado 1 no satisface las hipótesis de ninguno de estos teoremas. En efecto, si $(x, y) \neq (0, 0)$,

$$f''_{xy}(x, y) = f''_{yx}(x, y) = \frac{x^5 + 9x^4y^2 - 9x^2y^4 - y^6}{(x^2 + y^2)^3}$$

no es continua en $(0, 0)$ (véanse los límites según rectas $y = mx$). No se cumplen, pues, las hipótesis del teorema de Schwarz (y menos aún las del de Bonnet, como ya señalamos allí). Tampoco se cumplen las del teorema de Heffter-Young, ya que

$$f'_x(x, y) = y \frac{x^4 - y^4 + 4x^2y^2}{(x^2 + y^2)^2} \quad \text{si } (x, y) \neq (0, 0), \quad f'_x(0, 0) = 0$$

sí es continua, pero no es diferenciable en $(0, 0)$: en efecto, aunque existen $f''_{xx}(0, 0) = 0$, $f''_{xy}(0, 0) = -1$, no es cierto que

$$f'_x(x, y) - f'_x(0, 0) = -y + o\left(\sqrt{x^2 + y^2}\right)$$

pues ello equivaldría a

$$y \frac{x^4 - y^4 + 4x^2y^2}{(x^2 + y^2)^2} + y = 2yx^2 \frac{x^2 + 3y^2}{(x^2 + y^2)^2} = o\left(\sqrt{x^2 + y^2}\right)$$

Pero

$$\frac{1}{\sqrt{x^2 + y^2}} 2yx^2 \frac{x^2 + 3y^2}{(x^2 + y^2)^2} = 2yx^2 \frac{x^2 + 3y^2}{(x^2 + y^2)^{5/2}}$$

es una función homogénea de grado cero por lo que tiene en el origen límites distintos según cada recta $y = mx$, y por tanto no existe el límite en $(0, 0)$.

2. Distintas expresiones del término complementario

Recordemos primero la deducción de la fórmula de Taylor usando la integral definida, concretamente la regla de Barrow: de la igualdad

$$D_t[f(h-t)] = -f'(h-t),$$

deducimos

$$f(h) - f(0) = \int_0^h f'(h-t) dt$$

Integrando por partes ($du = dt, v = f'(h-t)$):

$$\begin{aligned} \int_0^h f'(h-t) dt &= t f'(h-t) \Big|_{t=0}^{t=h} + \int_0^h t f''(h-t) dt = \\ &= h f'(0) + \int_0^h t f''(h-t) dt \end{aligned}$$

Por tanto,

$$f(h) - f(0) - h f'(0) = \int_0^h t f''(h-t) dt$$

Volviendo a integrar por partes,

$$\begin{aligned} \int_0^h t f''(h-t) dt &= \frac{t^2}{2} f''(h-t) \Big|_{t=0}^{t=h} + \int_0^h \frac{t^2}{2} f'''(h-t) dt = \\ &= \frac{h^2}{2} f''(0) + \int_0^h \frac{t^2}{2} f'''(h-t) dt \end{aligned}$$

En general,

$$f(h) - f(0) - hf'(0) - \frac{h^2}{2}f''(0) - \dots - \frac{h^p}{p!}f^{(p)}(0) = \int_0^h \frac{t^p}{p!}f^{(p+1)}(h-t) dt$$

Haciendo ahora el cambio de variable $s = h - t$,

$$\int_0^h \frac{t^p}{p!}f^{(p+1)}(h-t) dt = \int_0^h \frac{(h-s)^p}{p!}f^{(p+1)}(s) ds$$

de donde se obtiene la fórmula de Taylor (o de Maclaurin)

$$f(h) = f(0) + hf'(0) + \frac{h^2}{2}f''(0) + \dots + \frac{h^p}{p!}f^{(p)}(0) + T_p(h)$$

con resto integral:

$$T_p(h) = \int_0^h \frac{(h-t)^p}{p!}f^{(p+1)}(t) dt$$

Para obtener las otras expresiones del término complementario que también se utilizan en la práctica, nos basamos en uno de los teoremas de la media del cálculo integral:

Sean $F, \phi: [a, b] \rightarrow \mathbb{R}$ continuas, $\phi > 0$ en (a, b) . Entonces existe $\tau \in (a, b)$ tal que

$$\int_a^b F(t)\phi(t) dt = F(\tau) \int_a^b \phi(t) dt$$

En efecto, si $m = \min F(t)$, $M = \max F(t)$, se tiene

$$m \int_a^b \phi(t) dt \leq \int_a^b F(t)\phi(t) dt \leq M \int_a^b \phi(t) dt$$

de donde, dado que $\int_a^b \phi(t) dt > 0$, resulta

$$m \leq \frac{\int_a^b F(t)\phi(t) dt}{\int_a^b \phi(t) dt} \leq M$$

y la continuidad de F implica la existencia de $\tau \in [a, b]$ tal que

$$F(\tau) = \frac{\int_a^b F(t)\phi(t) dt}{\int_a^b \phi(t) dt}$$

Un argumento algo más fino permite demostrar que siempre se puede elegir τ en el intervalo abierto (a, b) .

Pues bien, elijamos ahora $r \in [0, p]$ (no necesariamente entero), $\phi(t) = (h-t)^r$, $F(t) = \frac{(h-t)^{p-r}}{p!} f^{(p+1)}(t)$. Entonces, $\int_0^h (h-t)^r dt = \frac{h^{r+1}}{r+1}$, y la aplicación del teorema de la media antes demostrado nos da la existencia de $\tau \in [0, h]$ tal que

$$T_p(h) = \int_0^h \frac{(h-t)^p}{p!} f^{(p+1)}(t) dt = \frac{h^{r+1}}{r+1} \frac{(h-\tau)^{p-r}}{p!} f^{(p+1)}(\tau)$$

(término complementario de *Schlömilch*). Particularizando $r = 0$, se obtiene el término complementario de Cauchy

$$T_p(h) = h \frac{(h-\tau)^p}{p!} f^{(p+1)}(\tau)$$

y para $r = p$ se obtiene el término complementario de Lagrange

$$T_p(h) = \frac{h^{p+1}}{(p+1)!} f^{(p+1)}(\tau)$$

Todas estas fórmulas se aplican también al caso de que el punto en el que se evalúan las derivadas no sea 0 sino un x_0 arbitrario, así como al caso en que h sea negativo. Basta para ello efectuar el desarrollo para la función $g(t) \stackrel{\text{def}}{=} f(x_0 + th)$ en la forma

$$g(1) = g(0) + g'(0)h + \frac{1}{2!}g''(0) + \cdots + \frac{1}{p!}g^{(p)}(0) + T_p(1)$$

y observar que $g^{(k)}(t) = f^{(k)}(x_0 + th)h^k$.

Obtenemos así la versión general del desarrollo de Taylor:

$$f(x_0 + h) = f(x_0) + hf'(x_0) + \frac{h^2}{2}f''(x_0) + \cdots + \frac{h^p}{p!}f^{(p)}(x_0) + T_p(h)$$

donde:

$$T_p(h) = \int_0^1 \frac{(1-s)^p}{p!} f^{(p+1)}(x_0 + sh) h^{p+1} ds \quad (\text{forma integral}),$$

$$T_p(h) = \frac{h^{p+1}}{(p+1)!} f^{(p+1)}(x_0 + \theta h) \quad (\text{forma de Lagrange})$$

Si sólo se supone la continuidad de la derivada p -ésima y se efectúa el desarrollo de grado $p-1$ con término complementario de Lagrange:

$$\begin{aligned} f(x_0 + h) &= f(x_0) + hf'(x_0) + \frac{h^2}{2}f''(x_0) + \cdots + \frac{h^{p-1}}{(p-1)!}f^{(p-1)}(x_0) + \\ &\quad + \frac{h^p}{p!}f^{(p)}(x_0 + \theta h) \end{aligned}$$

se tiene $f^{(p)}(x_0 + \theta h) - f^{(p)}(x_0) \stackrel{\text{def}}{=} \delta(h) \rightarrow 0$ cuando $h \rightarrow 0$, y entonces

$$\begin{aligned} f(x_0 + h) &= f(x_0) + hf'(x_0) + \frac{h^2}{2}f''(x_0) + \cdots + \frac{h^{p-1}}{(p-1)!}f^{(p-1)}(x_0) + \\ &\quad + \frac{h^p}{p!}f^{(p)}(x_0) + \frac{h^p}{p!}\delta(h) \end{aligned}$$

de donde se obtiene el término complementario en forma infinitesimal, o asintótica, o de Peano:

$$T_p(h) = o(h^p)$$

que tiene la ventaja de no exigir la existencia de derivada de orden $p+1$. Si ésta existe y $|f^{(p+1)}(t)| \leq M$ en $[0, h]$, se puede mejorar esta expresión asintótica, y concluir que

$$|T_p(h)| \leq \frac{M}{(p+1)!}h^{p+1}$$

que es una forma más fuerte y "cuantitativa" que la mera expresión $T_p(h) = o(h^p)$.

Para una función f de n variables, el desarrollo de Taylor de la función de una variable $F(t) = f(x_0^0 + th_1, x_0^1 + th_2, \dots, x_0^{n-1} + th_n) = f(x_0 + th)$ en el intervalo $[0, 1]$ y la relación fundamental

$$F^{(k)}(t) = d^k f(x_0 + th; h) = \left(\sum_{i=1}^n h_i D_{x_i} \right)^k (f)(x_0 + th)$$

nos dan las fórmulas básicas del término complementario:

$$T_p(h) = \int_0^1 \frac{(1-s)^p}{p!} d^{p+1} f(x_0 + sh; h) ds \quad (\text{forma integral}),$$

$$T_p(h) = \frac{1}{(p+1)!} d^{(p+1)} f(x_0 + \theta h; h) \quad (\text{forma de Lagrange})$$

Si para dos funciones f y g definidas en un entorno de a introducimos el símbolo

$$f(x) = O(g(x)) \quad \text{para } x \rightarrow a$$

para indicar que $|f(x)/g(x)|$ permanece acotada en un entorno U de a , o, más precisamente, que existe una constante C tal que $|f(x)| \leq C|g(x)|$ para $x \in U$, $x \neq a$, (a puede ser $\pm\infty$, y f y g suelen ser infinitésimas o infinitas cuando $x \rightarrow a$), entonces podemos resumir de la siguiente forma las fórmulas

asintóticas del término complementario en función de las hipótesis sobre las derivadas de f :

$$\begin{aligned} f^{(p)} \text{ continua} &\Rightarrow T_p(h) = o(h^p) \\ f^{(p+1)} \text{ acotada} &\Rightarrow T_p(h) = O(h^{p+1}) \\ f^{(p+1)} \text{ continua} &\Rightarrow T_{p+1}(h) = o(h^{p+1}) \end{aligned}$$

En la práctica, la acotación de $f^{(p+1)}$ en $[0, h]$ es consecuencia de su continuidad, y entonces, si no se conoce una cota explícita para $f^{(p+1)}$ se puede usar la estimación asintótica fuerte $T_{p+1}(h) = o(h^{p+1})$ para el polinomio de Taylor de grado $p+1$. Si se conoce dicha cota (M), se podría usar la cota más precisa $|T_p(h)| \leq Mh^{p+1}/(p+1)!$ para el polinomio de Taylor de grado p . El usar una u otra información depende del problema que se esté analizando en cada caso.

Por ejemplo, anticipándonos al estudio general que se efectuará en el siguiente capítulo, si para una función de una variable $y = f(x)$ se tiene: $f'(x_0) = 0$, $f''(x_0) > 0$ y f'' es continua, entonces, usando la fórmula de Taylor de segundo grado en la forma asintótica $T_2(h) = o(h^2)$ se puede garantizar que en $x = x_0$ la función tiene un mínimo estricto local: existe un entorno U de x_0 tal que $f(x) > f(x_0)$ para todo $x \in U$, $x \neq x_0$.

Esta afirmación es de tipo cualitativo, y el tamaño de U no se puede estimar sin más información. Ésta puede venir dada por el conocimiento de una cota sobre f''' . Así, si $|f'''(x)| \leq M$ para todo x , tendríamos:

$$f(x_0 + h) = f(x_0) + \frac{f''(x_0)}{2}h^2 + \frac{f'''(\xi)}{6}h^3$$

donde ξ está entre 0 y h . Por tanto,

$$\left| f(x_0 + h) - f(x_0) - \frac{f''(x_0)}{2}h^2 \right| \leq \frac{M}{6}|h|^3$$

o sea

$$-\frac{M}{6}|h|^3 \leq f(x_0 + h) - f(x_0) - \frac{f''(x_0)}{2}h^2 \leq \frac{M}{6}|h|^3$$

de donde

$$f(x_0 + h) - f(x_0) \geq \frac{f''(x_0)}{2}h^2 + \frac{M}{6}|h|^3 = \frac{h^2}{2} \left(f''(x_0) + \frac{M}{3}|h| \right)$$

y de aquí concluimos que si $0 < |h| < \delta \stackrel{\text{def}}{=} 3f''(x_0)/M$, se tiene, efectivamente,

$$f(x_0 + h) > f(x_0)$$

3. El problema de la invariancia de las diferenciales de orden superior

Continuando con los comentarios de las notas 1 y 2 del apartado 3.4, comencemos recordando que, si $x = u(t)$; $y = v(t)$, la derivada primera de la función compuesta $F(t) \stackrel{\text{def}}{=} f(u(t), v(t))$ viene dada por

$$F'(t) = f'_x(u(t), v(t))u'(t) + f'_y(u(t), v(t))v'(t)$$

de donde

$$F'(t) = df(u(t), v(t); u'(t), v'(t))$$

Así, la expresión analítica de la diferencial $df = f'_x dx + f'_y dy$ se puede utilizar para recordar la fórmula anterior, ya que $dx = u'(t) dt$ y $dy = v'(t) dt$.

Recordemos ahora que, dada $f(x, y)$, se define

$$d^2 f(x_0, y_0; h, k) = f''_{xx}(x_0, y_0)h^2 + 2f''_{xy}(x_0, y_0)hk + f''_{yy}(x_0, y_0)k^2$$

Vimos también que la derivada segunda de $F(t)$ tiene por expresión

$$F''(t) = f''_{xx}u'(t)^2 + 2f''_{xy}u'(t)v'(t) + f''_{yy}v'(t)^2 + f'_x u''(t) + f'_y v''(t)$$

donde las derivadas de f están evaluadas en $(u(t), v(t))$. Obsérvese que esta expresión no coincide con $d^2 f(u(t), v(t); u'(t), v'(t))$, que sólo abarca a los tres primeros términos de la fórmula total. La expresión completa, escrita en términos diferenciales, sería:

$$F''(t) = d^2 f(u(t), v(t); u'(t), v'(t)) + df(u(t), v(t); u''(t), v''(t))$$

o, en forma equivalente,

$$\begin{aligned} d^2 F(t; h) &= d^2 f(u(t), v(t); dx(t; h), dy(t; h)) + \\ &+ df(u(t), v(t); d^2 x(t; h), d^2 y(t; h)) \end{aligned}$$

Esto justificaría llamar *expresión analítica de la segunda diferencial* a la siguiente:

$$d^2 z = z''_{xx} (dx)^2 + 2z''_{xy} (dx)(dy) + z''_{yy} (dy)^2 + z'_x d^2 x + z'_y d^2 y$$

para justificar la fórmula correcta (interpretando adecuadamente la notación):

$$\frac{d^2 z}{dt^2} = z''_{xx} \left(\frac{dx}{dt} \right)^2 + 2z''_{xy} \frac{dx}{dt} \frac{dy}{dt} + z''_{yy} \left(\frac{dy}{dt} \right)^2 + z'_x \frac{d^2 x}{dt^2} + z'_y \frac{d^2 y}{dt^2}$$

En esta forma, la expresión analítica de la segunda diferencial sd es la misma, tanto si x e y son variables independientes (en cuyo caso $d^2x = 0$, $d^2y = 0$) como si son funciones de otras variables s, w, \dots .

La derivada tercera de $F(t)$ viene dada por

$$F'''(t) = f_{x^3}'''(u')^3 + 3f_{x^2y}'''(u')^2v' + 3f_{xy^2}'''u'(v')^2 + f_{y^3}'''(v')^3 + 3f_{x^2}''u'u'' + 3f_{xy}''(u'v'' + u''v') + 3f_{y^2}''v'v'' + f_x'u''' + f_y'v'''$$

y la expresión analítica de la tercera diferencial sería

$$d^3z = f_{x^3}'''(dx)^3 + 3f_{x^2y}'''(dx)^2(dy) + 3f_{xy^2}'''(dx)(dy)^2 + f_{y^3}'''(dy)^3 + 3f_{x^2}''(dx)(d^2x) + 3f_{xy}''((dx)(d^2y) + (d^2x)(dy)) + 3f_{y^2}''(dy)(d^2y) + f_x'(d^3x) + f_y'(d^3y)$$

5.6 Problemas

1. Calcular las derivadas segundas y terceras de

$$a) f(x, y) = e^x + xy^2 \quad b) f(x, y) = x^2y^2 \quad c) f(x, y, z) = xyz$$

Clasificar la correspondientes matrices hessianas en el origen.

2. Estudiar la existencia y continuidad de las derivadas segundas de las funciones del problema 5 del capítulo 3. ¿Se verifica para ellas la igualdad de las derivadas cruzadas?
3. Obtener los desarrollos de Taylor de segundo orden de las siguientes funciones en el punto indicado:

$$a) f(x, y) = 1 - 2x + y - (x + y)^2, (0, 0)$$

$$b) f(x, y) = a + bx + cy + dx^2 + exy + ky^2, (0, 0)$$

$$c) f(x, y, z) = x - y^2 + xz, (1, 1, 1)$$

$$d) f(x, y) = \sqrt{1 - x^2 - y^2}, (0, 0)$$

$$e) f(x, y) = \cos(xy) + \sin(x^2 + y), (0, 0)$$

$$f) f(x, y) = e^x y, (0, 1)$$

$$g) f(x, y) = x^y, (1, 0)$$

$$h) f(x, y, z) = \sin(x + yz) - \ln(1 + xz), (0, 0, 0)$$

4. (Utilidad de la unicidad del polinomio de Taylor)

- (a) Obtener los desarrollos de Taylor de segundo orden en
- $(0,0)$
- y
- $(2,2)$
- de las funciones

$$(i) f(x, y) = e^{x-y} \qquad (ii) f(x, y) = \operatorname{sen}(x - y)$$

$$(iii) f(x, y) = \cos(x - y)$$

- (b) Teniendo en cuenta los desarrollos de Taylor de segundo orden de las siguientes funciones de una variables

$$e^z = 1 + z + \frac{z^2}{2} + \dots; \quad \operatorname{sen} z = z + \dots; \quad \cos z = 1 - \frac{z^2}{2} + \dots$$

sustituir $z = x - y$ en estos desarrollos y comparar con los obtenidos en a).

- (c) En general, dada una función
- $\varphi(z)$
- de una variable de clase
- C^2
- , obtener el desarrollo de Taylor de segundo orden de la función de dos variables
- $f(x, y) = \varphi(x - y)$
- en
- $(0,0)$
- (y en cualquier punto de la forma
- (a, a)
-), comprobando después que se puede obtener sustituyendo
- z
- por
- $x - y$
- en el desarrollo de Taylor de segundo orden de
- $\varphi(z)$
- en
- $z = 0$

$$\varphi(z) = \varphi(0) + \varphi'(0)z + \frac{1}{2}\varphi''(0)z^2 + \dots$$

5. Obtener los desarrollos de Taylor de cuarto orden de:

$$f(x, y) = \operatorname{sen}(xy); \quad f(x, y) = \operatorname{sen}(xyz); \quad f(x, y) = \operatorname{sen}(xy) \cos(x^2y)$$

empleando los desarrollos (de una variable) de las funciones seno y coseno.

6. Dada la función

$$z = x^2 + y^2x$$

se consideran valores determinados (x, y) a los cuales se dan incrementos $(\Delta x, \Delta y)$. Se pide:

- (a) Obtener exactamente el incremento
- Δz
- correspondiente.

- (b) Obtener aproximadamente Δu mediante las fórmulas de Taylor de primer y segundo grado, y comparar con el valor exacto de Δu .
- (c) Calcular numéricamente dichos valores para $x = 2, y = 1, \Delta x = 0,1, \Delta y = -0,2$. Lo mismo para $x = 2, y = 1, \Delta x = 0,001, \Delta y = -0,002$.

7. Se considera la ecuación

$$x^2 - y^2 - x + (x-1)z = 0 \quad (*)$$

- (a) Comprobar que $(*)$ define a z como función implícita de (x, y) en un entorno del punto $(0, 0, 0)$.
- (b) Obtener el polinomio de Taylor de segundo grado de z como función de (x, y) en el punto $(0, 0)$.
- (c) Comprobar que z se puede despejar efectivamente de $(*)$. Una vez obtenida la forma explícita de z , obtener su polinomio de Taylor en $(0, 0)$ y compararlo con el anteriormente obtenido.

8. Dada la función

$$f(x, y) = xy \ln(x^2 + y^2) \quad \text{si } (x, y) \neq (0, 0), \quad f(0, 0) = 0$$

se pide:

- (a) Escribir el polinomio de Taylor de segundo grado $P_2(x, y)$ de f en $(\sqrt{3}/2, 1/2)$ (véanse los problemas 5 del capítulo 3 y 2 del presente capítulo).
- (b) ¿Define la ecuación $f(x, y) = 0$ una función implícita $y = \varphi(x)$ en el entorno del punto $(\sqrt{3}/2, 1/2)$?
- (c) Si la respuesta a la pregunta anterior es afirmativa, determinar el polinomio de Taylor de segundo orden de $y = \varphi(x)$ en el entorno de $x_0 = \sqrt{3}/2$.
- (d) ¿Es posible describir la gráfica de $y = \varphi(x)$ en términos geométricos simples?

Ecuaciones en derivadas parciales Como en el problema 15 del capítulo 3, una *ecuación en derivadas parciales* es una relación que ha de satisfacer una "función incógnita" junto con sus derivadas parciales hasta un determinado orden. El orden de una ecuación es el de la derivada de mayor orden que aparece en ella. Toda función que, sustituida —a la vez que sus derivadas— en una ecuación, convierte a ésta en una identidad en las variables independientes,

se dice que es una *solución* de la ecuación o que *satisface* la ecuación. Las ecuaciones en derivadas parciales más importantes en las aplicaciones son, sin duda, las de segundo orden. Entre ellas destacan las ecuaciones fundamentales de la física matemática:

$$\frac{\partial^2 u}{\partial t^2} - c^2 \frac{\partial^2 u}{\partial x^2} = 0 \quad \text{ecuación de ondas}$$

$$\frac{\partial u}{\partial t} - k^2 \frac{\partial^2 u}{\partial x^2} = 0 \quad \text{ecuación del calor}$$

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0 \quad \text{ecuación de Laplace}$$

(t representa al tiempo y x, y, \dots variables espaciales; en el segundo miembro puede haber una función distinta de la trivial, y entonces la ecuación no es *homogénea*).

9. Obtener el conjunto de soluciones de las ecuaciones en derivadas parciales:

$$a) u_{xx} = 0 \quad b) u_{yy} = 0 \quad c) u_{xy} = 0 \quad d) u_{xy} = x^2 y$$

(Indicación: Se entiende que se buscan las funciones de clase C^2 en todo el plano que verifican cada una de las ecuaciones. Véase el apartado 4.1. En el primer caso, por ejemplo, de $(u_x)_x = 0$ se deduce que $u_x(x, y) = \varphi(y)$, e, integrando de nuevo respecto a x , se tiene $u(x, y) = x\varphi(y) + \psi(y)$, siendo φ y ψ funciones arbitrarias. En c), la solución general es $u(x, y) = \varphi(x) + \psi(y)$ con φ y ψ funciones arbitrarias de una variable.)

10. Se trata de integrar la ecuación de ondas

$$\frac{\partial^2 u}{\partial t^2} - c^2 \frac{\partial^2 u}{\partial x^2} = 0 \quad (*)$$

(o sea, de obtener el conjunto de sus soluciones). Esta ecuación rige las oscilaciones planas de una cuerda que se supone de longitud infinita y cuya posición de equilibrio coincide con el eje x ; $u(x, t)$ representa el desplazamiento vertical respecto al eje de abscisas del punto x en el instante t . Para un t fijado la gráfica de $x \mapsto u(x, t)$ representa el perfil de la cuerda en ese instante.

- (a) Transformar la ecuación (*) mediante un cambio de variables *lineal*

$$x = A\xi + B\eta, \quad y = C\xi + D\eta$$

Elegir los coeficientes A, B, C, D de modo que la nueva ecuación en las variables ξ, η sea

$$\frac{\partial^2 u}{\partial \xi \partial \eta} = 0$$

- (b) Utilizar c) del problema anterior para obtener que las soluciones de (*) están dadas por

$$u(x, t) = \varphi(x + ct) + \psi(x - ct) \quad (**)$$

donde φ y ψ son funciones arbitrarias de una variable.

- (c) Se trata ahora de probar que entre la infinidad de soluciones (**) hay una sola que satisface las condiciones iniciales

$$u(x, 0) = f(x); \quad u_t(x, 0) = g(x) \quad (***)$$

donde $f(x)$ y $g(x)$ son dos funciones dadas de una variable que representan, respectivamente, la *posición inicial* y la *velocidad inicial* de la cuerda (se tendrá, por tanto, que si se prefijan una posición y una velocidad iniciales, quedará bien determinado un único movimiento de la cuerda). Impónganse las condiciones (***) en la expresión (**) para determinar cómo han de ser φ y ψ . La solución buscada es

$$u(x, t) = \frac{1}{2} [f(x + ct) + f(x - ct)] + \frac{1}{2c} \int_{x-ct}^{x+ct} g(s) ds$$

- (d) Resolver el problema de valor inicial

$$\begin{cases} u_{tt} - 4u_{xx} = 0 \\ u(x, 0) = \cos x \\ u_t(x, 0) = \sin x \end{cases}$$

11. Si $u(x, y)$ ($u(x, y, z)$) es de clase C^2 en un conjunto abierto Ω , la función

$$\Delta u = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} \quad \left(\Delta u = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} \right)$$

se llama *laplaciana* de la función u . Una función cuya laplaciana es nula en Ω (o sea, que satisface la ecuación de Laplace en Ω) se dice que es *armónica* en Ω . (La definición es la misma, naturalmente, para cualquier número de variables.)

- (a) Comprobar que son armónicas las siguientes funciones, precisando el mayor abierto Ω en que ello es así:

$$(i) x^2 - y^2 \quad (ii) x^3 - 3xy^2 + xy \quad (iii) \arctan \frac{y}{x}$$

- (b) Dada la función de tres variables

$$r(x, y, z) = \sqrt{(x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2}$$

calcúlese Δr y $\Delta \frac{1}{r}$. ¿Qué se obtiene para la función de n variables, $n > 3$,

$$r(x_1, \dots, x_n) = \sqrt{(x_1 - x_1^0)^2 + (x_2 - x_2^0)^2 + \dots + (x_n - x_n^0)^2} ?$$

- (c) La misma cuestión para la función de dos variables

$$r(x, y) = \sqrt{(x - x_0)^2 + (y - y_0)^2}$$

Calcúlese también en este caso $\Delta \left(\ln \frac{1}{r} \right) = \Delta (-\ln r)$.

- (d) Si $u(x, y) = f(r)$, con $r(x, y) = \sqrt{x^2 + y^2}$, expresar Δu en función de r . ¿Cómo ha de ser f para que se tenga $\Delta u = 0$? Resolver la misma cuestión para tres variables (x, y, z) y comentar los resultados obtenidos en relación con los dos apartados anteriores de este problema.
- (e) Comprobar que la expresión de la laplaciana de $u(x, y)$ en coordenadas polares es

$$\Delta u = u_{rr} + \frac{1}{r} u_r + \frac{1}{r^2} u_{\theta\theta}$$

y que son armónicas en $R^2 \setminus \{(0, 0)\}$ las funciones

$$(i) r^n \cos n\theta \quad (ii) r^n \sin n\theta$$

(Estas funciones armónicas en $R^n \setminus \{0\}$ de los puntos b), c), d) y e) juegan un papel esencial en la resolución de problemas relativos a la ecuación de Laplace.)

12. La ecuación del calor

$$\frac{\partial u}{\partial t} - k^2 \frac{\partial^2 u}{\partial x^2} = 0$$

rige la distribución de temperaturas $u(x, t)$ a lo largo del tiempo —se sobreentiende, en general, que, en la ecuación, $t > 0$ — en un medio unidimensional (una barra metálica, por ejemplo) que se supone ocupando el eje de abscisas (o un intervalo de este eje). (También se la denomina *ecuación de la difusión* porque rige diversos fenómenos de difusión en la física y otras ciencias.) Probar que las siguientes funciones son soluciones de la ecuación del calor:

$$\begin{array}{ll} a) e^{-k^2 m^2 t} \cos mx & b) e^{-k^2 m^2 t} \sin mx \quad (m \text{ constante}) \\ c) \frac{1}{\sqrt{4\pi t}} \exp\left(-\frac{x^2}{4k^2 t}\right) & \end{array}$$

(Gracias a estas soluciones especiales se resuelven muchos problemas relativos a la ecuación del calor, como se muestra en los cursos de ecuaciones en derivadas parciales.)

13. Determinar cuáles de los siguientes campos vectoriales son campos gradiente, hallando una función potencial para aquéllos que lo sean:

$$\begin{array}{ll} (i) (y - \ln x, x + e^y) & (ii) (\cos x + y + \sin y, \cos y + x + x \cos y) \\ (iii) (1 + e^{xy}, 1 + e^{xy}) & (iv) \left(\frac{y^2 + 2xy + ax^2}{(x^2 + y^2)^2}, -\frac{x^2 + 2xy + bx^2}{(x^2 + y^2)^2} \right) \\ & \text{(según los valores de las constantes } a \text{ y } b) \\ (v) \text{ (tres variables)} & \left(\frac{1}{x^2 y} - \frac{1}{x^2 + z^2}, \frac{z}{xy^2}, \frac{x}{x^2 + z^2} - \frac{1}{xy} \right) \end{array}$$

14. Una ecuación diferencial $x' = f(t, x)$ siempre se puede escribir en la forma

$$m(t, x) + n(t, x) \frac{dx}{dt} = 0 \quad (*)$$

En efecto, basta tomar $n(t, x) \equiv 1$, $m(t, x) = -f(t, x)$. Caben otras elecciones: si, por ejemplo, $f(t, x)$ es de la forma

$$f(t, x) = -\frac{f_1(t, x)}{f_2(t, x)}$$

entonces la ecuación puede escribirse como

$$f_1(t, x) + f_2(t, x) \frac{dx}{dt} = 0$$

La ecuación diferencial (*) se dice que es *exacta* si $m_x = n_t$, o sea, si (m, n) es un campo gradiente. Cuando esto ocurre, las soluciones de la ecuación están definidas implícitamente por la ecuación

$$F(t, x) = C \quad (**)$$

donde F es una función potencial del campo (m, n) y C cualquier constante. En efecto, si $x(t)$ está definida por (**) (obsérvese que hay realmente ecuación diferencial (*) si $n(t, x) = F_x(t, x) \neq 0$, que es precisamente la hipótesis del teorema de la función implícita que garantiza que (**) define a x como función de t en un intervalo I), entonces, por la regla de la cadena,

$$F_t(t, x(t)) + F_x(t, x(t))x'(t) = 0$$

es decir, $x(t)$ satisface la ecuación diferencial (*). Recíprocamente, si $x(t)$ es solución de (*) en un cierto intervalo I , entonces

$$\begin{aligned} 0 &= m(t, x(t)) + n(t, x(t)) \frac{dx}{dt} = \frac{\partial F}{\partial t}(t, x(t)) + \frac{\partial F}{\partial x}(t, x(t)) \cdot \frac{dx}{dt} = \\ &= \frac{d}{dt}(F(t, x(t))) = 0 \text{ para todo } t \in I \end{aligned}$$

es decir, $F(t, x(t)) = \text{constante}$ para todo $t \in I$, lo que quiere decir que $x(t)$ está definida por la ecuación (**) para un cierto valor de la constante C .

(La expresión (*) de una ecuación diferencial es la manera "moderna" —sustituyendo diferenciales por derivadas— de escribir lo que más tradicionalmente se expresa en la llamada "forma diferencial"

$$m(t, x) dt + n(t, x) dx = 0$$

En esta forma, una ecuación diferencial —*ecuación entre diferenciales*— es exacta si el primer miembro de la ecuación es la diferencial —también en escritura tradicional— de una función de dos variables $F(t, x)$.

Comprobar que las siguientes ecuaciones son exactas y encontrar su *solución general*:

a) $x^2 + (2tx + \operatorname{sen} x)x' = 0$

b) $x(2t - x) + (t^2 - 2tx)x' = 0$

c) $x(2t - x) + (t^2 - 2tx)x' = 0$

Determinar la solución que satisface la condición inicial $x(-1) = 1$

d) $\frac{2t}{x^3} + \frac{x^2 - 3t^2}{x^4} x' = 0$

Determinar la solución que satisface la condición inicial $x(1) = 1$

15. Puede ocurrir que la ecuación

$$m(t, x) + n(t, x) \frac{dx}{dt} = 0$$

no sea exacta pero que sí lo sea la ecuación equivalente que resulta al multiplicarla por una función no nula $\mu(t, x)$:

$$\mu(t, x)m(t, x) + \mu(t, x)n(t, x) \frac{dx}{dt} = 0$$

La condición para ello es que

$$\frac{\partial}{\partial x}(\mu n) = \frac{\partial}{\partial t}(\mu m)$$

es decir

$$\mu_x m + \mu m_x = \mu_t n + \mu n_t$$

Una función $\mu(t, x)$ que satisfaga esta ecuación se dice que es un *factor integrante* de la ecuación diferencial dada. En general, la determinación de un factor integrante es un problema más difícil que la resolución de la propia ecuación original. Sólo en ciertos casos, por la naturaleza de la ecuación o por el trasfondo físico de la misma, es viable conjeturar y comprobar que ciertas funciones simples son factores integrantes.

Resolver las siguientes ecuaciones diferenciales utilizando un factor integrante del tipo que se indica:

a) $2x^4 + t^3 + 2tx^3x' = 0$, $\mu = t^3$

b) $x + 2t + (xt + t + t^2)x' = 0$, $\mu = e^x$

c) $x + (1 + xe^{-t})x' = 0$, μ función sólo de t : $\mu = \mu(t)$

d) $x(t+1) + t(x+1)x' = 0$, $\mu = \mu(x+t)$

e) $6x^2 + 6xt + (9tx + 4t^2)x' = 0$, $\mu = \mu(xt)$

16. Probar que toda ecuación de variables separables, o sea, de la forma

$$m(t) + n(x)x' = 0$$

es exacta.

17. (a) Una ecuación de la forma

$$x' = a(t)x + b(t) \quad (*)$$

donde $a(t)$ y $b(t)$ son funciones continuas en un intervalo $(\alpha, \omega) \subseteq \mathbb{R}$, se dice que es **lineal** (por ser el segundo miembro una expresión lineal en la función incógnita x). Probar que la ecuación $(*)$ admite un factor integrante que es sólo función de t y escribir la fórmula que da su solución general.

- (b) Resolver los siguientes problemas de valor inicial:

$$(i) \begin{cases} x' + 2tx = t^3 \\ x(0) = 1 \end{cases} \quad (ii) \begin{cases} x' + tx = t \\ x(0) = 1 \end{cases} \quad (iii) \begin{cases} x' - 2tx = 1 \\ x(0) = 1 \end{cases}$$

18. La ecuación de Bernoulli es de la forma

$$x' = a(t)x + b(t)x^n$$

- (a) Probar que el cambio de variable $u = x^{1-n}$ la convierte en una ecuación lineal en la nueva variable dependiente u . Probar también que la ecuación de Bernoulli admite un factor integrante de la forma $\mu(t, x) = A(t)B(x)$ y obtenerlo.

- (b) Resolver

$$(i) x' = \frac{4t^3 - x^3}{3tx^2} \quad (ii) \begin{cases} 4x^3x' + \frac{x^4}{t} = 6 \\ x(1) = 2 \end{cases}$$

19. La ecuación de Riccati tiene la forma

$$x' = p(t)x^2 + q(t)x + r(t) \quad (*)$$

- (a) Probar que si se conoce una solución particular $x_1(t)$, entonces el cambio de variable dado por

$$x = x_1(t) + \frac{1}{v}$$

donde v es la nueva variable dependiente, reduce $(*)$ a una ecuación lineal. (Equivalentemente, el cambio de variable $v = x - x_1(t)$ la reduce a una ecuación de Bernoulli.)

- (b) Aplicar lo anterior a la resolución de las ecuaciones:

$$(i) x' = \frac{t+1}{2t^2}x^2 + \frac{1}{2t}x - \frac{t}{2} \quad (ii) x' = x^2 - e^tx + e^t$$

$$(\text{solución particular } x_1(t) = t) \quad (\text{solución particular } x_1(t) = e^t)$$

Capítulo 6

Optimización

En el apartado 1.3 se exponían las primeras ideas sobre los problemas de optimización, o sea, de determinación de máximos y mínimos. Veíamos allí algunas situaciones especiales en las que era posible dicha determinación por procedimientos gráficos, sin mucho aparato teórico. Ahora ya disponemos de los conceptos necesarios para desarrollar las bases teóricas de las estrategias para resolver problemas de optimización. Sin proponernos realizar un estudio exhaustivo, veremos, sucesivamente, los problemas sin restricciones, con restricciones de igualdad y con restricciones de desigualdad, con una atención especial a los problemas convezos, de gran importancia en las aplicaciones y que admiten un tratamiento más completo.

6.1 Máximos y mínimos

Repitamos aquí las definiciones dadas en el capítulo 1:

Definición 6.1 Sean $f : \text{Dom } f \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ y un conjunto $S \subseteq \text{Dom } f$. Se dice que la función f alcanza su máximo en S en el punto $x_0 \in S$ si $f(x) \leq f(x_0)$ para todo $x \in S$. $f(x_0)$ se dice que es el valor máximo de f en S . Si $f(x) < f(x_0)$ para todo $x \in S$ distinto de x_0 se dice que f tiene un máximo estricto en x_0 . Se definen de manera análoga mínimo y mínimo estricto cambiando el signo de las desigualdades.

El término de **óptimo** se refiere indistintamente a máximo y mínimo. Con cierto abuso de lenguaje se dirá coloquialmente que el punto x_0 donde se alcanza un óptimo es un máximo (o un mínimo) de la función f .

Un resultado teórico fundamental en la determinación de máximos y mínimos es el siguiente:

Teorema 6.1 (Teorema de Weierstrass) Si la función $f : \text{Dom } f \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ es continua en un conjunto cerrado y acotado $S \subseteq \text{Dom } f$, entonces f alcanza en S su valor máximo y su valor mínimo.

Este teorema, que demostraremos en el capítulo 7 que sigue, generaliza el correspondiente resultado para funciones de una variable continuas en un intervalo cerrado y acotado $[a, b]$. En la figura 6.1 se muestran diversas situaciones en una dimensión que ilustran la influencia de las hipótesis del teorema. Discútanse como ejercicio.

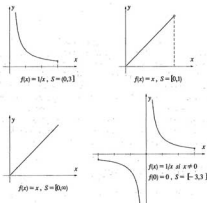


Figura 6.1

El teorema de Weierstrass garantiza la existencia de óptimo en un conjunto cerrado y acotado pero *no* proporciona ningún procedimiento para localizar en la práctica los puntos en que lo alcanza. Es en este punto donde comienza a ser útil el cálculo diferencial:

Teorema 6.2 Supongamos que f alcanza su máximo (o su mínimo) en $S \subseteq \text{Dom } f$ en un punto interior de S x_0 . Si las derivadas parciales $\partial f / \partial x_i$, $i = 1, \dots, n$, existen en x_0 entonces

$$\frac{\partial f}{\partial x_i}(x_0) = 0, \quad i = 1, \dots, n \quad (6.1)$$

En efecto, como x_0 es punto interior, existe $\delta > 0$ tal que $x_0 + te_i \in S$ para $|t| < \delta$ y basta aplicar el resultado análogo para funciones de una variable a cada función parcial

$$x_i \mapsto f(x_1^0, \dots, x_{i-1}^0, x_i, \dots, x_n^0)$$

la cual tiene en el intervalo $(x_i^0 - \delta, x_i^0 + \delta)$ un máximo (resp. mínimo) en x_i^0 .

La condición (6.1) es necesaria ("condición necesaria de primer orden") pero no suficiente para la existencia de óptimo. Por ejemplo, para la función $f(x, y) = y^2 - x^2$, definida en todo \mathbb{R}^2 , el único punto del plano que verifica

$$\frac{\partial f}{\partial x}(x_0, y_0) = 0, \quad \frac{\partial f}{\partial y}(x_0, y_0) = 0$$

es el origen $(0, 0)$. Pero $f(x, 0) = -x^2$, de donde, con $x \neq 0$, se tiene $f(x, 0) < f(0, 0) = 0$ y, por tanto, no puede haber un mínimo en $(0, 0)$. Por otro lado, $f(0, y) = y^2 > 0$ para $y \neq 0$, con lo que tampoco puede haber un máximo. Recordemos que esto también se daba en funciones de una variable: la función $f(x) = x^2$ verifica $f'(0) = 0$ y, sin embargo no tiene un máximo ni un mínimo en $x_0 = 0$.

Definición 6.2 Un punto $x_0 \in \text{Int}(\text{Dom } f)$ se dice que es un **punto crítico** de f si

$$\frac{\partial f}{\partial x_1}(x_0) = \dots = \frac{\partial f}{\partial x_n}(x_0) = 0 \quad (6.2)$$

Con esta definición se tiene el siguiente

Corolario 6.1 Supongamos que f tiene derivadas parciales en un conjunto abierto $D \supseteq S$. Si f alcanza en S su valor máximo (o su valor mínimo) en el punto $x_0 \in S$, entonces o bien x_0 es un punto crítico de f o bien $x_0 \in \partial S$.

EJEMPLO 1. Funciones lineales. Para una función lineal

$$f(x) = a_1x_1 + \dots + a_nx_n + b$$

se tiene

$$\frac{\partial f}{\partial x_1} = a_1, \dots, \frac{\partial f}{\partial x_n} = a_n$$

y, por tanto, no tiene puntos críticos salvo en el caso trivial de las funciones constantes $f(x) = b$, para las que todo punto $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ es punto

crítico. En consecuencia, las funciones lineales no pueden alcanzar su valor óptimo —máximo o mínimo— en el interior de un conjunto S ; si lo alcanza ha de ser en la frontera de S .

Sobre la base de este corolario se desarrolla la estrategia para localizar máximos y mínimos análoga a la que conocemos de cursos anteriores para funciones de una variable: si se trata de optimizar $f(x)$ en un intervalo cerrado y acotado $[a, b]$, sabemos de antemano, por el teorema de Weierstrass, que el problema tiene solución; se comienza por resolver la ecuación $f'(x) = 0$; supongamos que x_1, \dots, x_N son las soluciones —puntos críticos de f — que pertenecen al intervalo (a, b) —que es el interior de $[a, b]$ —. Se comparan los valores $f(a), f(x_1), \dots, f(x_N), f(b)$, y el mayor de ellos proporciona el máximo y el menor el mínimo (Fig. 6.2. Si el óptimo se realiza en uno de los extremos, no tiene por qué anularse en él la derivada: la condición necesaria del teorema es para puntos interiores).

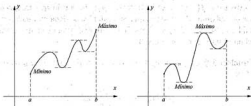


Figura 6.2

Si se trata de hallar los valores máximo y mínimo de una función en un intervalo I abierto en alguno de sus extremos (en particular, en todo \mathbb{R}) habrá que comparar los valores de f en los puntos x_1, \dots, x_N tales que $f'(x_j) = 0$, $j = 1, \dots, N$, con $\lim f(x)$ cuando x tiende a dicho extremo (eventualmente, $x \rightarrow \infty$ y/o $x \rightarrow -\infty$). Esto ya es, en general, más complicado y no hay garantía a priori de que la función alcanza efectivamente en I un valor máximo y/o un valor mínimo (Fig. 6.3).

Veamos algunos ejemplos de funciones de dos variables.

EJEMPLO 2. Determinar los valores máximo y mínimo de $f(x, y) = xy$ en el conjunto $B_1(0) = \{(x, y); x^2 + y^2 \leq 1\}$. El conjunto $S = B_1(0)$ es cerrado y acotado, con lo que, por el teorema de Weierstrass, $f(x, y) = xy$ alcanza en él su valor máximo y su valor mínimo. Lo hará en el interior o en la frontera. Si lo hace en el interior será en un punto crítico, o sea, una solución del sistema

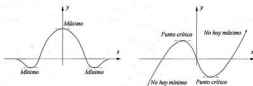


Figura 6.3

de puntos críticos

$$\begin{aligned}\frac{\partial f}{\partial x} &= y = 0 \\ \frac{\partial f}{\partial y} &= x = 0\end{aligned}$$

Vemos que el único punto crítico es el $(0, 0)$. La frontera es la circunferencia de ecuación $x^2 + y^2 = 1$. Para evaluar f sobre la circunferencia, escribamos las ecuaciones paramétricas de ésta:

$$\begin{cases} x = \cos \theta \\ y = \sin \theta \end{cases}, \quad 0 \leq \theta \leq 2\pi$$

Entonces

$$f(\cos \theta, \sin \theta) = \frac{1}{2} \sin 2\theta \stackrel{\text{def}}{=} F(\theta)$$

Los puntos críticos de $F(\theta)$ en $[0, 2\pi]$ satisfacen $F'(\theta) = 0$, $0 \leq \theta \leq 2\pi$, es decir, son $\theta = \frac{\pi}{4}, \frac{3\pi}{4}, \frac{5\pi}{4}, \frac{7\pi}{4}$. Hemos de considerar también los puntos extremos $\theta = 0$ y $\theta = 2\pi$. Tenemos así cinco puntos "candidatos" a óptimos en la frontera del conjunto:

$$A\left(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\right); B\left(-\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\right); C\left(-\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2}\right); D\left(\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2}\right); E(1, 0)$$

Se tiene

$$f(A) = \frac{1}{2}; f(B) = -\frac{1}{2}; f(C) = \frac{1}{2}; f(D) = -\frac{1}{2}; f(E) = 0$$

Se compara con el valor en el origen, el "candidato" del interior de S , $f(0, 0) = 0$, y vemos que el máximo tiene lugar en A y C y el mínimo en B y D (Fig. 6.4).

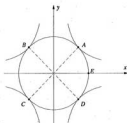


Figura 6.4

Cuando la frontera sea de estructura compleja la resolución puede ser muy difícil (en una variable, la frontera de un intervalo cerrado y acotado consiste simplemente en los dos puntos extremos; en dos y más dimensiones la frontera de un conjunto puede ser un conjunto de estructura complicada). Por otra parte, un problema como el anterior se resuelve más fácilmente, convenientemente reinterpretado, por el método de los multiplicadores de Lagrange que veremos en el apartado 6.2.

EJEMPLO 3. Sea $f(x, y) = -x^2 + x - y^2 - 2y - 1$; $S = \mathbb{R}^2$.

El sistema de puntos críticos es

$$\begin{aligned}\frac{\partial f}{\partial x} &= -2x + 1 = 0 \\ \frac{\partial f}{\partial y} &= -2y - 2 = 0\end{aligned}$$

con lo que el único punto crítico es $(x_0, y_0) = (\frac{1}{2}, -1)$. $f(\frac{1}{2}, -1) = \frac{1}{4}$. Completando cuadrados:

$$f(x, y) = -\left(x - \frac{1}{2}\right)^2 - (y + 1)^2 + \frac{1}{4}$$

vedmos que $f(x, y) \leq \frac{1}{4}$ para todo $(x, y) \in \mathbb{R}^2$ por lo que $\max_{\mathbb{R}^2} f(x, y) = \frac{1}{4}$ y se alcanza en $(\frac{1}{2}, -1)$. En este ejemplo, un simple argumento algebraico nos da que la función alcanza su valor máximo (estricto, ya que $f(x, y) < \frac{1}{4} = f(\frac{1}{2}, -1)$ si $(x, y) \neq (\frac{1}{2}, -1)$) aunque no podemos aplicar el teorema de Weierstrass por no tratarse de un conjunto acotado. Por otro lado, f no alcanza su valor mínimo en \mathbb{R}^2 (¿por qué?).

En general, si el conjunto no es cerrado y acotado, habrá que estudiar el comportamiento de f al “acercarse a la frontera”, en particular “en el infinito” si S no es acotado (es difícil precisar más estas expresiones; cada caso, si es tratable analíticamente, sugiere el método de hacerlo).

EJEMPLO 4. Determinar los valores máximo y mínimo en \mathbb{R}^2 de la función

$$f(x, y) = (1 - x^2 - y^2)e^{-x^2 - y^2}$$

Puntos críticos:

$$\begin{aligned}\frac{\partial f}{\partial x} &= 2x(x^2 + y^2 - 2)e^{-x^2 - y^2} = 0 \\ \frac{\partial f}{\partial y} &= 2y(x^2 + y^2 - 2)e^{-x^2 - y^2} = 0\end{aligned}$$

sistema satisfecho por los puntos

$$(0, 0) \text{ y } (\bar{x}, \bar{y}) \text{ tales que } \bar{x}^2 + \bar{y}^2 = 2$$

Se tiene $f(0, 0) = 1$, $f(\bar{x}, \bar{y}) = -1/e^2$. Por otro lado

$$f(x, y) = (1 - r^2)e^{-r^2} \stackrel{\text{def}}{=} \varphi(r), \text{ con } r = \sqrt{x^2 + y^2}$$

Como $\lim_{r \rightarrow \infty} \varphi(r) = 0$, por la regla de L'Hôpital (la gráfica de $\varphi(r)$ es la de la figura 6.3a), el máximo se alcanza en $(0, 0)$ y el mínimo en los puntos de la circunferencia $x^2 + y^2 = 2$.

Ya se comprende que estos métodos *ad hoc* no llegan muy lejos en cuanto la función y/o el conjunto S se complican mínimamente. Por un lado, vamos a profundizar en el análisis de los puntos críticos —en tanto que “candidatos” a óptimos— y, por otro, en relación a la influencia de la geometría del conjunto S en que se plantea el problema de optimización, vamos a considerar una familia especial de problemas de optimización —los problemas convexos— que se presentan con frecuencia en las aplicaciones y para los cuales puede llevarse a cabo un tratamiento muy completo.

Óptimos locales. Clasificación de puntos críticos

En la figura 6.2 observamos varios puntos críticos de funciones de una variable en los que no se alcanza el máximo o el mínimo de la definición 1. Son máximos y mínimos *locales*:

Definición 6.3 Sean $f : \text{Dom } f \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ y $S \subseteq \text{Dom } f$. Se dice que la función f tiene un **máximo local** (relativamente a S) en el punto $x_0 \in S$ si existe $\varepsilon > 0$ tal que $f(x) \leq f(x_0)$ para todo $x \in B_\varepsilon(x_0) \cap S$. Si $f(x) < f(x_0)$ para todo $x \in B_\varepsilon(x_0) \cap S$, $x \neq x_0$, se dice que f tiene un **máximo local estricto** en x_0 . Se definen de manera análoga **mínimo local** y **mínimo local estricto**.

Los máximos y mínimos de la definición 1 son los óptimos **globales** de f en S . Es obvio que un máximo (mínimo) global también lo es local. Por el mismo argumento que antes se tiene:

Teorema 6.3 Supongamos que f tiene derivadas parciales en un conjunto abierto $D \supseteq S$. Si f tiene un óptimo local en un punto x_0 del interior de S , entonces x_0 es un punto crítico de f .

En la figura 6.5 se ilustran distintas posibilidades para funciones de una

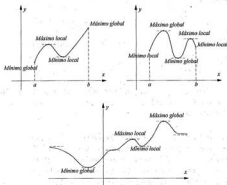


Figura 6.5

variable que, con gráficas más complicadas de representar, se dan también para funciones de dos variables (Fig. 6.6).

Los puntos críticos de una función f —que, por definición, son puntos interiores de $\text{Dom } f$ — se clasifican en

- **máximos locales:** existe $\varepsilon > 0$ tal que $f(x) \leq f(x_0)$ para todo $x \in B_\varepsilon(x_0)$.

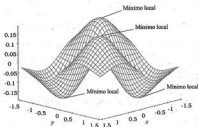


Figura 6.6

- **mínimos locales:** existe $\varepsilon > 0$ tal que $f(x) \geq f(x_0)$ para todo $x \in B_\varepsilon(x_0)$.
- **puntos de silla:** en toda bola $B_\varepsilon(x_0)$ centrada en x_0 existen puntos de $\text{Dom } f$ en los que $f(x) < f(x_0)$ y puntos de $\text{Dom } f$ en los que $f(x) > f(x_0)$.

El punto $(1/2, -1)$ del ejemplo 3 anterior es un **máximo** (global, de hecho). Los puntos críticos de $f(x, y) = x^2 + y^2$ son las soluciones del sistema

$$f_x(x, y) = 2x = 0$$

$$f_y(x, y) = 2y = 0$$

es decir, $(0, 0)$ es el único punto crítico. Como $f(x, y) = x^2 + y^2 > f(0, 0) = 0$ para todo $(x, y) \neq (0, 0)$, se trata de un **mínimo global estricto**. Vimos también que $(0, 0)$ es el único punto crítico de $f(x, y) = y^2 - x^2$ y que no es ni máximo ni mínimo, es decir que se trata de un **punto de silla** o **puerto** (Fig. 38 del capítulo 1; es el “genuino” punto de silla o puerto, por recordar, como decíamos en el capítulo 1, la forma de una silla de montar o un puerto de montaña; en otros casos, la gráfica de la función $f(x, y)$ no tiene por qué tener ese aspecto en el entorno de (x_0, y_0) , pero se sigue utilizando esa denominación para un punto crítico que no es ni máximo ni mínimo. En una dimensión, $x_0 = 0$ es un ejemplo para $f(x) = x^3$ de punto crítico que no es ni máximo ni mínimo; esbócese la gráfica de la función de dos variables $f(x, y) = x^3$).

Nos interesa establecer criterios que permitan decidir a qué tipo pertenece un punto crítico. Ello nos permitirá, por ejemplo, descartar los puntos de silla

y los mínimos locales en un problema de maximización.

Condiciones de segundo orden para óptimos locales

Son condiciones basadas en las derivadas segundas (de ahí lo de "segundo orden") que generalizan el **criterio de la derivada segunda** para funciones de una variable que conocemos de cursos anteriores: dado un punto crítico x_0 de la función $f(x)$

- Si $f''(x_0) < 0$, entonces f tiene un máximo local (estricto) en x_0 .
- Si $f''(x_0) > 0$, entonces f tiene un mínimo local (estricto) en x_0 .
- Si f tiene en x_0 un máximo local, entonces $f''(x_0) \leq 0$.
- Si f tiene en x_0 un mínimo local, entonces $f''(x_0) \geq 0$.

Para obtener un criterio análogo para funciones de dos y más variables partimos del desarrollo de Taylor de segundo orden:

$$\begin{aligned} f(x) - f(x_0) &= [f'_{x_1}(x_0)h_1 + \dots + f'_{x_n}(x_0)h_n] + \\ &+ \frac{1}{2} \sum_{i,j=1}^n f''_{x_i x_j}(x_0)h_i h_j + R_2(h) \end{aligned} \quad (6.3)$$

donde $h = (h_1, \dots, h_n) = x - x_0$ y $R_2(h) = o(|h|^2)$. Se supone que f es de clase C^2 en un entorno del punto crítico x_0 . Por ser x_0 punto crítico los n primeros sumandos del segundo miembro de (6.3) valen cero, con lo que

$$f(x) - f(x_0) = \frac{1}{2} d^2 f(x_0; h) + R_2(h) \quad (6.4)$$

Para clasificar el punto crítico x_0 hay que estudiar el signo de $[f(x) - f(x_0)]$ para puntos x próximos a x_0 , es decir, el signo de $\frac{1}{2} d^2 f(x_0; h) + R_2(h)$ para $|h|$ pequeño. Pero, dado que $R_2(h)$ es pequeño para $|h|$ pequeño, es razonable conjeturar que, en un entorno de $h = 0$, el signo de $\frac{1}{2} d^2 f(x_0; h) + R_2(h)$ venga determinado por el de $\frac{1}{2} d^2 f(x_0; h)$.

$\frac{1}{2} d^2 f(x_0; h)$ es un polinomio homogéneo de grado 2 en $h = (h_1, \dots, h_n)$, es decir, una *forma cuadrática*. Una forma cuadrática de n variables tiene por expresión

$$q(h) = \sum_{i,j=1}^n a_{ij} h_i h_j, \quad a_{ij} = a_{ji}, \quad h \in \mathbb{R}^n \quad (6.5)$$

o, utilizando notación matricial,

$$q(h) = (h_1, h_2, \dots, h_n) \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} \begin{pmatrix} h_1 \\ h_2 \\ \vdots \\ h_n \end{pmatrix} = h^T A h$$

donde $A = (a_{ij})$, $i, j = 1, \dots, n$, la matriz de coeficientes, es simétrica: $a_{ij} = a_{ji}$ (los términos de cualquier polinomio homogéneo de segundo grado $\sum_{i,j=1}^n b_{ij} h_i h_j$ se pueden reagrupar poniendo $a_{ij} = a_{ji} = \frac{1}{2}(b_{ij} + b_{ji})$ para que resulte una expresión (6.5) con coeficientes simétricos; en otras palabras, la matriz de coeficientes de uno de tales polinomios siempre se puede suponer simétrica).

Los valores (y, por tanto, el signo) de una forma cuadrática están determinados, gracias a la homogeneidad, por los valores que toma en un entorno del origen. En efecto, sea $\delta > 0$ y sea $h \in \mathbb{R}^n$, $h \neq 0$. El punto $\frac{\delta}{|h|}h$ está en la recta que pasa por el origen y el punto h , y tiene norma δ .

Por ser $q(h)$ homogénea de grado 2 se tiene

$$q\left(\frac{\delta}{|h|}h\right) = \frac{\delta^2}{|h|^2}q(h)$$

y, por tanto,

$$q(h) = \begin{cases} \frac{|h|^2}{\delta^2} q\left(\frac{\delta}{|h|}h\right) & \text{si } h \neq 0 \\ 0 & \text{si } h = 0 \end{cases} \quad (6.6)$$

es decir, $q(h)$ queda determinada por los valores que toma en la esfera $\{h \in \mathbb{R}^n; |h| = \delta\}$ cualquiera que sea el $\delta > 0$ elegido y su signo es el mismo para todos los puntos de cada rayo que parte del origen. En consecuencia, estudiar el signo de $d^2f(x_0; h)$ para $|h|$ pequeño es equivalente a estudiarlo cuando $h = (h_1, \dots, h_n)$ recorre \mathbb{R}^n .

De acuerdo con los signos que toman en los puntos de \mathbb{R}^n , las formas cuadráticas (o las matrices simétricas que las definen) se clasifican en los siguientes tipos:

1. Semidefinida positiva si $q(h) \geq 0$ para todo $h \in \mathbb{R}^n$.
2. Definida positiva si $q(h) > 0$ para todo $h \in \mathbb{R}^n$, $h \neq 0$.
3. Semidefinida negativa si $q(h) \leq 0$ para todo $h \in \mathbb{R}^n$.
4. Definida negativa si $q(h) < 0$ para todo $h \in \mathbb{R}^n$, $h \neq 0$.

5. **Indefinida** si existen h^* y h^{**} tales que $q(h^*) > 0$ y $q(h^{**}) < 0$.

NOTA. El polinomio cero, $q(h) \equiv 0$, será simultáneamente semidefinido positivo y semidefinido negativo.

Pues bien, como respuesta a la conjetura formulada anteriormente se tiene el siguiente criterio que, efectivamente, generaliza el criterio de la derivada segunda para funciones de una variable:

Teorema 6.4 Sea f de clase C^2 en un conjunto abierto S y sea $x_0 \in S$ un punto crítico de f . Entonces:

1. Si $d^2f(x_0; h)$ es definida positiva, x_0 es un mínimo local estricto.
2. Si $d^2f(x_0; h)$ es definida negativa, x_0 es un máximo local estricto.
3. Si $d^2f(x_0; h)$ es indefinida, x_0 es un punto de silla.
4. Si x_0 es un mínimo local, $d^2f(x_0; h)$ es semidefinida positiva.
5. Si x_0 es un máximo local, $d^2f(x_0; h)$ es semidefinida negativa.

NOTAS. 1. Si $n = 1$, $d^2f(x_0; h) = f''(x_0)h^2$; para $n = 1$ no cabe la posibilidad de que una forma cuadrática sea indefinida; no tiene, pues, sentido, si $n = 1$, el punto 3 del criterio.

2. Los criterios 1, 2 y 3 son **condiciones suficientes** para que un punto crítico sea del tipo que se indica, mientras que 4 y 5 son **condiciones necesarias**. Obsérvese que el teorema no agota todas las posibilidades: cuando $d^2f(x_0; h)$ es semidefinida positiva pero no positiva o semidefinida negativa pero no negativa se trata de un caso dudoso y habría que recurrir a términos del desarrollo de Taylor de orden superior al segundo.

EJEMPLO 5.

1. $f(x, y) = x^2 + y^3$; $(0, 0)$ es el único punto crítico. $f''_{xx}(0, 0) = 2$, $f''_{yy}(0, 0) = 0$, $f''_{xy}(0, 0) = 0$; $d^2f((0, 0); (h_1, h_2)) = 2h_1^2$ es semidefinida positiva (en los puntos $(0, h_2)$, $h_2 \neq 0$, vale 0) y $(0, 0)$ es un punto de silla (¿por qué?).
2. $f(x, y) = x^2 + y^4$; $(0, 0)$ es el único punto crítico y $d^2f((0, 0); (h_1, h_2))$ es la misma que antes, pero ahora $(0, 0)$ es un mínimo global estricto.
3. Háganse análogas consideraciones para las funciones $f(x, y) = -x^2 + y^3$, $f(x, y) = -x^2 - y^4$.

Demostraremos este teorema más adelante. Para su utilización en las aplicaciones es esencial disponer de criterios prácticos que permitan clasificar las formas cuadráticas según los tipos descritos anteriormente, cuestión que se estudia en los cursos de álgebra lineal. Repasaremos aquí algunos aspectos de la misma.

Dos variables

Una forma cuadrática de dos variables es un polinomio

$$q(h, k) = ah^2 + 2bhk + ck^2 \quad (6.7)$$

que, utilizando notación matricial, podemos escribir en la forma

$$q(h, k) = (h, k) \begin{pmatrix} a & b \\ b & c \end{pmatrix} \begin{pmatrix} h \\ k \end{pmatrix} \quad (6.8)$$

La clasificación de (6.7) en términos de los coeficientes a , b y c es algebraicamente muy simple: completando cuadrados se tiene, supuesto que $a \neq 0$:

$$\begin{aligned} q(h, k) &= ah^2 + 2bhk + ck^2 + \frac{b^2}{a}k^2 - \frac{b^2}{a}k^2 = \\ &= a \left(h^2 + 2\frac{b}{a}hk + \frac{b^2}{a^2}k^2 \right) + \left(c - \frac{b^2}{a} \right) k^2 = \\ &= a \left(h + \frac{b}{a}k \right)^2 + \frac{ac - b^2}{a}k^2 \end{aligned}$$

Por tanto:

$$ah^2 + 2bhk + ck^2 = \begin{cases} a \left(h + \frac{b}{a}k \right)^2 + \frac{ac - b^2}{a}k^2 & \text{si } a \neq 0 \\ k^2 (2b\frac{b}{k} + c) & \text{si } a = 0, k \neq 0 \\ 0 & \text{si } a = 0, k = 0 \end{cases}$$

de donde se deduce inmediatamente el siguiente criterio:

$$\begin{aligned} q(h, k) \text{ es definida positiva} &\iff a > 0, \begin{vmatrix} a & b \\ b & c \end{vmatrix} > 0 \\ q(h, k) \text{ es definida negativa} &\iff a < 0, \begin{vmatrix} a & b \\ b & c \end{vmatrix} > 0 \\ q(h, k) \text{ es semidefinida positiva} &\iff a \geq 0, c \geq 0, \begin{vmatrix} a & b \\ b & c \end{vmatrix} \geq 0 \\ q(h, k) \text{ es semidefinida negativa} &\iff a \leq 0, c \leq 0, \begin{vmatrix} a & b \\ b & c \end{vmatrix} \geq 0 \\ q(h, k) \text{ es indefinida} &\iff \begin{vmatrix} a & b \\ b & c \end{vmatrix} < 0 \end{aligned} \quad (6.9)$$

(Obsérvese que las condiciones $a > 0$, $ac - b^2 > 0$ llevan implícito el que también $c > 0$; y, análogamente, $a < 0$, $ac - b^2 > 0$ el que $c < 0$.) Del teorema 6.4 se deriva entonces el siguiente criterio para clasificar puntos críticos de funciones de dos variables:

Teorema 6.5 Sea $f(x, y)$ de clase C^2 en un conjunto abierto $S \subseteq \mathbb{R}^2$ y sea $(x_0, y_0) \in S$ un punto crítico de f . Entonces:

- Si $f''_{xx}(x_0, y_0) > 0$ y $\begin{vmatrix} f''_{xx}(x_0, y_0) & f''_{xy}(x_0, y_0) \\ f''_{xy}(x_0, y_0) & f''_{yy}(x_0, y_0) \end{vmatrix} > 0$, f tiene en (x_0, y_0) un mínimo local estricto.
- Si $f''_{xx}(x_0, y_0) < 0$ y $\begin{vmatrix} f''_{xx}(x_0, y_0) & f''_{xy}(x_0, y_0) \\ f''_{xy}(x_0, y_0) & f''_{yy}(x_0, y_0) \end{vmatrix} > 0$, f tiene en (x_0, y_0) un máximo local estricto.
- Si $\begin{vmatrix} f''_{xx}(x_0, y_0) & f''_{xy}(x_0, y_0) \\ f''_{xy}(x_0, y_0) & f''_{yy}(x_0, y_0) \end{vmatrix} < 0$, f tiene en (x_0, y_0) un punto de silla.
- Finalmente, si se verifica que $\begin{vmatrix} f''_{xx}(x_0, y_0) & f''_{xy}(x_0, y_0) \\ f''_{xy}(x_0, y_0) & f''_{yy}(x_0, y_0) \end{vmatrix} = 0$, o sea, $f''_{xx}(x_0, y_0)f''_{yy}(x_0, y_0) - f''_{xy}(x_0, y_0)^2 = 0$, (x_0, y_0) puede ser un mínimo local, un máximo local o un punto de silla (caso dudoso).

EJEMPLO 6. Determinar y clasificar los puntos críticos de la función $f(x, y) = x^3 + 3xy^2 - 3x^2 - 3y^2 + 4$.

El sistema de puntos críticos es

$$\begin{cases} f_x(x, y) = 3x^2 + 3y^2 - 6x = 0 \\ f_y(x, y) = 6xy - 6y = 0 \end{cases}$$

y sus soluciones

$$P_1(0, 0), P_2(2, 0), P_3(1, 1), P_4(1, -1)$$

La matriz hessiana de $f(x, y)$ es

$$D^2f(x, y) = \begin{pmatrix} f''_{xx}(x, y) & f''_{xy}(x, y) \\ f''_{xy}(x, y) & f''_{yy}(x, y) \end{pmatrix} = \begin{pmatrix} 6x - 6 & 6y \\ 6y & 6x - 6 \end{pmatrix}$$

La evaluamos en cada uno de los puntos críticos:

$$\begin{aligned} D^2f(0, 0) &= \begin{pmatrix} -6 & 0 \\ 0 & -6 \end{pmatrix}; \quad D^2f(2, 0) = \begin{pmatrix} 6 & 0 \\ 0 & 6 \end{pmatrix}; \\ D^2f(1, 1) &= \begin{pmatrix} 0 & 6 \\ 6 & 0 \end{pmatrix}; \quad D^2f(1, -1) = \begin{pmatrix} 0 & -6 \\ -6 & 0 \end{pmatrix} \end{aligned}$$

y vemos que:

- $P_1(0, 0)$ es un máximo local (estricto),

- $P_2(2, 0)$ es un mínimo local (estricto),
- $P_3(1, 1)$ es un punto de silla, y
- $P_4(1, -1)$ es un punto de silla.

Consideremos el caso especial en que $f(x, y)$ es una forma cuadrática:

$$f(x, y) = ax^2 + 2bxy + cy^2 \quad (6.10)$$

Sus puntos críticos son las soluciones de

$$\begin{cases} f_x(x, y) = 2ax + 2by = 0 \\ f_y(x, y) = 2bx + 2cy = 0 \end{cases} \quad (6.11)$$

La matriz de coeficientes del sistema lineal homogéneo (6.11) coincide con la matriz hessiana de f

$$D^2 f(x, y) = 2 \begin{pmatrix} a & b \\ b & c \end{pmatrix}$$

y es la matriz de coeficientes de la forma cuadrática multiplicada por 2. Si $\begin{vmatrix} a & b \\ b & c \end{vmatrix} \neq 0$, $(0, 0)$ es la única solución del sistema (6.11), o sea, el único punto crítico de f . Si $\begin{vmatrix} a & b \\ b & c \end{vmatrix} > 0$ se tratará de un óptimo global, por la propia definición de forma cuadrática definida positiva (en el caso de un mínimo) o negativa (en el caso de un máximo). Si $\begin{vmatrix} a & b \\ b & c \end{vmatrix} = 0$, el sistema (6.11) tiene infinitas soluciones y la función f tiene toda una recta de puntos críticos (supuesto que no se trata del caso trivial $a = b = c = 0$, o sea, $f(x, y) \equiv 0$, en el que todos los puntos del plano son críticos). Todos ellos serán mínimos globales (no estrictos) si $f(x, y)$ es semidefinida positiva o bien máximos globales (no estrictos) si $f(x, y)$ es semidefinida negativa. Obsérvese que cuando $f(x, y)$ es ella misma una forma cuadrática los óptimos son globales y que en el caso dudoso del teorema 4 no se da la posibilidad de punto de silla.

EJERCICIO. Analizar los puntos críticos de un polinomio general de segundo grado

$$f(x, y) = ax^2 + 2bxy + cy^2 + dx + ey + f$$

La intuición ligada a la representación gráfica que se trae de las funciones de una variable puede inducir a error en funciones de dos y más variables. Por ejemplo, en una variable, una función (derivable) que tiene en un punto un mínimo local y que no tiene más puntos críticos, alcanza en él su mínimo global. En dos variables, sin embargo, el resultado análogo no es cierto: la función

$$f(x, y) = x^2(1 + y)^3 + y^2$$

por ejemplo, tiene a $(0, 0)$ como único punto crítico, es un mínimo local y f no alcanza en él su valor mínimo en \mathbb{R}^2 . En una variable, una función no puede tener, por ejemplo, sólo dos puntos críticos que sean ambos mínimos, pero la función de dos variables

$$f(x, y) = (x^2y - x - 1)^2 + (x^2 - 1)^2$$

por ejemplo, tiene en $(1, 2)$ y $(-1, 0)$ dos mínimos locales y ningún otro punto crítico. (Compruébense estas afirmaciones.)

n variables

Para $n > 2$, la obtención de un criterio de clasificación como el dado en (6.9) es, naturalmente, más complicado. Si $n = 3$, todavía son relativamente fáciles manipulaciones algebraicas mediante el artificio de completar cuadrados. Supuesto que

$$D_1 = a_{11} \neq 0 \text{ y } D_2 = \begin{vmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}^2 \neq 0$$

se tiene

$$\begin{aligned} q(h) &= a_{11}h_1^2 + 2a_{12}h_1h_2 + 2a_{13}h_1h_3 + 2a_{23}h_2h_3 + a_{22}h_2^2 + a_{33}h_3^2 = \\ &= a_{11} \left(h_1 + \frac{a_{12}}{a_{11}}h_2 + \frac{a_{13}}{a_{11}}h_3 \right)^2 + \frac{D_2}{a_{11}} \left(h_2 + \frac{a_{11}a_{23} - a_{12}a_{13}}{D_2}h_3 \right)^2 + \frac{D_3}{D_2}h_3^2 \end{aligned} \quad (6.12)$$

donde $D_3 = |A| = a_{11}a_{22}a_{33} - a_{11}a_{23}^2 - a_{22}a_{13}^2 - a_{33}a_{12}^2 + 2a_{12}a_{13}a_{23}$.

Al quedar $q(h)$ escrita como suma de términos cuadráticos, su signo dependerá del signo de los coeficientes

$$a_{11} = D_1, \quad \frac{D_2}{D_1}, \quad \frac{D_3}{D_2}$$

Así pues,

1. $q(h) > 0$ para todo $h \in \mathbb{R}^3$, $h \neq 0$, o sea, $q(h)$ es definida positiva si y sólo si $D_1 > 0$, $\frac{D_2}{D_1} > 0$, $\frac{D_3}{D_2} > 0$ lo que equivale a que

$$D_1 > 0, D_2 > 0, D_3 > 0 \quad (6.13)$$

((6.13) también es condición necesaria pues $q(h) \neq 0$ para todo $h \in \mathbb{R}^3$, $h \neq 0$, implica que $a_{11} \neq 0$, $D_2 \neq 0$, y valen los cálculos que llevan a (6.13)).

2. $q(h) < 0$ para todo $h \in \mathbb{R}^3$, $h \neq 0$, o sea, $q(h)$ es definida negativa si y sólo si $D_1 < 0$, $\frac{D_2}{D_1} < 0$, $\frac{D_3}{D_2} < 0$ lo que equivale a que

$$D_1 < 0, D_2 > 0, D_3 < 0 \quad (6.14)$$

(vale la misma observación que en el punto anterior).

- Si $D_1 > 0$, $D_2 > 0$, $D_3 = 0$, $q(h)$ es semidefinida positiva.
- Si $D_1 < 0$, $D_2 > 0$, $D_3 = 0$, $q(h)$ es semidefinida negativa.
- Si D_1, D_2, D_3 son todos distintos de cero y no se dan ni 1) ni 2) o bien $D_1 \neq 0$, $D_2 \neq 0$, $D_3 = 0$ y no se dan ni 3) ni 4), entonces $q(h)$ es indefinida.

De lo anterior se deriva el siguiente criterio (simplificado) para clasificar puntos críticos de funciones de 3 variables:

Teorema 6.6 Sea $f(x, y, z)$ de clase C^2 en un conjunto abierto $S \subseteq \mathbb{R}^3$ y sea $(x_0, y_0, z_0) \in S$ un punto crítico de f . Sean

$$H_1 = f''_{xx}(x_0, y_0, z_0), \quad H_2 = \begin{vmatrix} f''_{xx} & f''_{xy} \\ f''_{xy} & f''_{yy} \end{vmatrix}_{(x_0, y_0, z_0)},$$

$$H_3 = \begin{vmatrix} f''_{xx} & f''_{xy} & f''_{xz} \\ f''_{yx} & f''_{yy} & f''_{yz} \\ f''_{zx} & f''_{zy} & f''_{zz} \end{vmatrix}_{(x_0, y_0, z_0)}$$

y supongamos que ninguno de los números H_1, H_2, H_3 es cero. Entonces

- Si $H_1 > 0$, $H_2 > 0$, $H_3 > 0$, f tiene un mínimo local (estricto) en (x_0, y_0, z_0) .
- Si $H_1 < 0$, $H_2 > 0$, $H_3 < 0$, f tiene un máximo local (estricto) en (x_0, y_0, z_0) .

3. En los seis casos restantes de sucesión de signos de H_1, H_2, H_3 , f tiene en (x_0, y_0, z_0) un punto de silla.

EJEMPLO 7. Determinar y clasificar los puntos críticos de la función

$$f(x, y, z) = x^3 + y^2 + z^2 - 2xy - x - 2z$$

Sistema de puntos críticos:

$$f_x(x, y, z) = 3x^2 - 2y - 1 = 0$$

$$f_y(x, y, z) = 2y - 2x = 0$$

$$f_z(x, y, z) = 2z - 2 = 0$$

Resolviéndolo se obtienen dos puntos críticos:

$$P_1(1, 1, 1) \text{ y } P_2(-1/3, -1/3, 1)$$

La matriz hessiana es

$$\begin{pmatrix} f''_{xx} & f''_{xy} & f''_{xz} \\ f''_{yx} & f''_{yy} & f''_{yz} \\ f''_{zx} & f''_{zy} & f''_{zz} \end{pmatrix} = \begin{pmatrix} 6x & -2 & 0 \\ -2 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}$$

Entonces:

$$H_1(P_1) = 6 > 0,$$

$$H_2(P_1) = \begin{vmatrix} 6 & -2 \\ -2 & 2 \end{vmatrix} = 8 > 0,$$

$$H_3(P_1) = 2 \cdot 8 = 16 > 0$$

$$H_1(P_2) = -2 < 0,$$

$$H_2(P_2) = \begin{vmatrix} -2 & -2 \\ -2 & 2 \end{vmatrix} = -8 < 0,$$

$$H_3(P_2) = 2 \cdot (-8) = -16 < 0$$

Por lo tanto, $P_1(1, 1, 1)$ es un mínimo local y $P_2(-1/3, -1/3, 1)$ un punto de silla.

En general, para una forma cuadrática de n variables $q(h) = \sum_{i,j=1}^n a_{ij} h_i h_j = h^T A h$, con $A = (a_{ij})$, $a_{ij} = a_{ji}$, sean

$$D_k = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ a_{21} & a_{22} & \cdots & a_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{k1} & a_{k2} & \cdots & a_{kk} \end{vmatrix}, \quad k = 1, \dots, n$$

los *menores principales dominantes* de la matriz $A = (a_{ij})$ de coeficientes. Entonces, se puede demostrar (véanse [2] y [8]) el siguiente resultado:

1. $q(h)$ es definida positiva si y sólo si $D_k > 0$ para todo $k = 1, 2, \dots, n$.
2. $q(h)$ es definida negativa si y sólo si $(-1)^k D_k > 0$ para todo $k = 1, 2, \dots, n$.
3. Si $D_k > 0$, $k = 1, 2, \dots, n-1$, $D_n = 0$, entonces $q(h)$ es semidefinida positiva.
4. Si $(-1)^k D_k > 0$, $k = 1, 2, \dots, n-1$, $D_n = 0$, entonces $q(h)$ es semidefinida negativa.
5. Si $D_n \neq 0$ y no se dan ni 1) ni 2) o bien si $D_k \neq 0$, $k = 1, 2, \dots, n-1$ y $D_n = 0$ no dándose ni 3) ni 4), entonces $q(h)$ es indefinida.

Podría pensarse que una forma cuadrática es semidefinida si y sólo si en 1) y 2) se sustituyen todas las desigualdades estrictas por desigualdades no estrictas \geq ó \leq . Pero no es así. Por ejemplo,

$$q(h_1, h_2, h_3) = -h_2^2 - h_3^2 = 0 \cdot h_1^2 + 0 \cdot h_1 h_2 + 0 \cdot h_1 h_3 + 0 \cdot h_2 h_3 - h_2^2 - h_3^2$$

es semidefinida negativa, no semidefinida positiva, a pesar de que los menores principales dominantes de la matriz de coeficientes

$$\begin{pmatrix} 0 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix}$$

son todos ≥ 0 (iguales a cero, de hecho).

Para establecer una condición necesaria y suficiente para que $q(h)$ sea semidefinida hay que hacer intervenir todos los menores principales de la matriz A de coeficientes. Un *menor principal* de orden k de A es el determinante que se obtiene suprimiendo $n - k$ filas y las $n - k$ columnas de los mismos índices. Se puede demostrar que una forma cuadrática $h^T A h$ es semidefinida positiva si y sólo si todos los menores principales de A son ≥ 0 y que es semidefinida negativa si y sólo si todos los menores principales de orden k , $k = 1, \dots, n$, de A o son nulos o tienen el mismo signo que $(-1)^k$ (esto se cumple en el ejemplo anterior. Compruébese, por otra parte, que este criterio es, para $n = 2$, el que vimos anteriormente).

También se pueden clasificar las formas cuadráticas en términos de los *autovalores* de la matriz A de coeficientes. Veamos: en los cursos de álgebra lineal (véase también el final del apartado 6.2 que sigue) se demuestra que, por

ser A simétrica, existe una matriz ortogonal P (o sea, tal que $P^{-1} = P^T$) tal que

$$B = P^T A P = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & \\ \dots & \dots & \dots & \\ 0 & & \dots & \lambda_n \end{pmatrix} \quad (6.15)$$

siendo $\lambda_1, \dots, \lambda_n$ los autovalores —reales— de A (o sea, las raíces de la ecuación característica $\det(A - \lambda I) = 0$). Se tiene entonces que

$$\begin{aligned} q(h) &= \sum_{i,j=1}^n a_{ij} h_i h_j = h^T A h = \\ &= \text{(haciendo el cambio de variables en } \mathbb{R}^n \text{ } h = Py) = \\ &= (Py)^T A (Py) = y^T (P^T A P) y = y^T B y = \\ &= \lambda_1 y_1^2 + \lambda_2 y_2^2 + \dots + \lambda_n y_n^2 \end{aligned} \quad (6.16)$$

de donde se deduce inmediatamente que

1. $q(h)$ es definida positiva \iff todos los autovalores de A son positivos.
2. $q(h)$ es semidefinida positiva \iff todos los autovalores de A son ≥ 0 .
3. $q(h)$ es definida negativa \iff todos los autovalores de A son negativos.
4. $q(h)$ es semidefinida negativa \iff todos los autovalores de A son ≤ 0 .
5. $q(h)$ es indefinida $\iff A$ tiene al menos dos autovalores con signos opuestos.

Este criterio es exhaustivo y muy fácil de establecer, pero tiene, frente al anterior basado en los menores principales —que puede suponer cálculos engorrosos pero siempre realizables—, la dificultad práctica de que para dimensiones altas ($n > 2$, de hecho) puede ser difícil —si no imposible— calcular los autovalores. Bien es verdad que, en realidad, lo que se necesita es el **signo** de los mismos y no su valor exacto, y hoy existen procedimientos numéricos fácilmente realizables en ordenador para obtener esa información.

Demostración del Teorema 6.4

1.- Según vimos, en el entorno de un punto crítico x_0 se tiene

$$f(x) - f(x_0) = \frac{1}{2} d^2 f(x_0; h) + R_2(h) \quad (6.17)$$

luego, para demostrar que x_0 es un mínimo local estricto bajo la hipótesis de que $\frac{1}{2}d^2f(x_0; h)$ es definida positiva, bastará probar que existe $\varepsilon > 0$ tal que

$$\frac{\frac{1}{2}d^2f(x_0; h) + R_2(h)}{|h|^2} > 0 \quad (6.18)$$

siempre que $0 < |h| < \varepsilon$. Aplicando la expresión (6.6)

$$q(h) = \begin{cases} \frac{|h|^2}{\delta^2} q\left(\frac{\delta}{|h|}h\right) & \text{si } h \neq 0 \\ 0 & \text{si } h = 0 \end{cases}$$

con $\delta = 1$ a la forma cuadrática $\frac{1}{2}d^2f(x_0; h)$, se tendrá para $|h| \neq 0$

$$\frac{\frac{1}{2}d^2f(x_0; h)}{|h|^2} = \frac{1}{2}d^2f(x_0; \frac{h}{|h|}) \quad (6.19)$$

$\frac{1}{2}d^2f(x_0; h)$ es una función continua en $h = (h_1, \dots, h_n)$, $\frac{h}{|h|} \in S^{n-1} = \{h \in \mathbb{R}^n; |h| = 1\}$ y S^{n-1} es un conjunto cerrado y acotado de \mathbb{R}^n , con lo que, por el teorema de Weierstrass, $\frac{1}{2}d^2f(x_0; h)$ alcanza en S^{n-1} un valor mínimo m , y $m > 0$ ya que, por hipótesis, $\frac{1}{2}d^2f(x_0; h) > 0$ para todo $h \neq 0$. En consecuencia

$$\frac{\frac{1}{2}d^2f(x_0; h) + R_2(h)}{|h|^2} \geq m + \frac{R_2(h)}{|h|^2} \quad (6.20)$$

Puesto que

$$\lim_{|h| \rightarrow 0} \frac{R_2(h)}{|h|^2} = 0 \quad (6.21)$$

existe $\varepsilon > 0$ tal que, si $0 < |h| < \varepsilon$,

$$\frac{|R_2(h)|}{|h|^2} < \frac{m}{2}$$

de donde, por (6.20),

$$\frac{\frac{1}{2}d^2f(x_0; h) + R_2(h)}{|h|^2} \geq m - \frac{m}{2} > 0$$

como queríamos demostrar.

4.- Supongamos ahora que f tiene un mínimo en x_0 , o sea, que existe $\varepsilon > 0$ tal que $f(x) \geq f(x_0)$ para todo $x \in B_\varepsilon(x_0) \cap S$. Como S es abierto, podemos suponer que $B_\varepsilon(x_0) \subset S$. Por (6.17) se tendrá que

$$\frac{1}{2}d^2f(x_0; h) + R_2(h) \geq 0 \quad (6.22)$$

para todo h tal que $|h| < \varepsilon$. Supongamos que existe $\bar{h} \in \mathbb{R}^n$ tal que $d^2f(x_0; \bar{h}) < 0$. Se trata de ver que, entonces, existen puntos de la semirrecta $\{\lambda\bar{h}; \lambda > 0\}$ con $|\lambda\bar{h}| < \varepsilon$ en los que

$$\frac{1}{2}d^2f(x_0; \lambda\bar{h}) + R_2(\lambda\bar{h}) < 0$$

lo que supone una contradicción con (6.22). Se tiene, por ser $d^2f(x_0; \lambda\bar{h})$ homogénea de grado 2,

$$\frac{\frac{1}{2}d^2f(x_0; \lambda\bar{h}) + R_2(\lambda\bar{h})}{|\lambda\bar{h}|^2} = \frac{1}{2} \frac{d^2f(x_0; \bar{h})}{|\bar{h}|^2} + \frac{R_2(\lambda\bar{h})}{|\lambda\bar{h}|^2} = -\alpha + \frac{R_2(\lambda\bar{h})}{|\lambda\bar{h}|^2}$$

con $\alpha > 0$. Como antes, por (6.21), existe δ , $0 < \delta \leq \varepsilon$, tal que

$$\frac{|R_2(\lambda\bar{h})|}{|\lambda\bar{h}|^2} < \frac{\alpha}{2}$$

si $0 < |\lambda\bar{h}| < \delta$. Tomando $\lambda > 0$ suficientemente pequeño para que $|\lambda\bar{h}| < \delta$ se tendrá

$$\frac{\frac{1}{2}d^2f(x_0; \lambda\bar{h}) + R_2(\lambda\bar{h})}{|\lambda\bar{h}|^2} < -\alpha + \frac{\alpha}{2} < 0$$

Ha de ser, por tanto, $d^2f(x_0; h) \geq 0$ para todo $h \in \mathbb{R}^n$.

Los puntos 2) y 5) del teorema se demuestran de manera análoga a 1) y 4).

3.- Si $d^2f(x_0; h)$ es indefinida, se toman \bar{h} y h^* fijos tales que $d^2f(x_0; \bar{h}) < 0$ y $d^2f(x_0; h^*) > 0$. Procediendo como en la demostración de 2), se prueba que en cualquier entorno $B_\varepsilon(x_0)$ hay puntos $x = x_0 + \lambda\bar{h}$ en los que $f(x) < f(x_0)$ y puntos $x_0 + \mu h^*$ en los que $f(x) > f(x_0)$, es decir, que x_0 es un punto de silla.

Problemas convexos

Como decíamos, si se dan determinadas propiedades de convexidad, el problema de la búsqueda de óptimos se simplifica notablemente. Concretamente, los óptimos locales —detectados por las condiciones necesarias de primer orden— serán *globales* (que son los óptimos que normalmente interesan en las aplicaciones) y dichas condiciones pasarán a ser también suficientes.

Para funciones de una variable (y también para las de dos variables) se puede dar una definición de *función conveza* en términos geométricos muy intuitivos: una función f definida en un intervalo $I \subseteq \mathbb{R}$ se dice que es **conveza** en I si para todo par de puntos $P_1(x_1, f(x_1))$, $P_2(x_2, f(x_2))$ de su gráfica, el segmento $[P_1, P_2]$ está situado en o por encima de dicha gráfica (o sea, para todo $x \in [x_1, x_2]$ la ordenada $f(x)$ del punto $(x, f(x))$ es menor o igual que la del punto de la recta P_1P_2 que tiene por abscisa x). Esos segmentos que unen pares de puntos de la gráfica de una función se llaman *cuerdas* de la función.

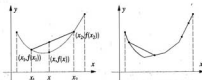


Figura 6.7

En el ejemplo de la figura 6.7a, todas las cuerdas están (excepto, naturalmente, los puntos extremos) por encima de la gráfica de la función. Cuando esto ocurre se dice que la función es *estrictamente conveza*.

De modo análogo, se dice que una función $f(x)$ es **cóncava** en I si para todo par de puntos de su gráfica, la cuerda que los une está en o por debajo de dicha gráfica. La función es estrictamente cóncava si todas las cuerdas de su gráfica están por debajo de ésta (equivalentemente, f es (estrictamente) cóncava si $-f$ es (estrictamente) convexa).

Las funciones convexas (cóncavas) se pueden caracterizar algebraicamente mediante el siguiente resultado:

Proposición 6.1 *La función f es conveza en el intervalo I si y sólo si cualesquiera que sean $x_1, x_2 \in I$ y $t \in [0, 1]$ se tiene que*

$$f((1-t)x_1 + tx_2) \leq (1-t)f(x_1) + tf(x_2) \quad (6.23)$$

(equivalentemente

$$f(\alpha_1 x_1 + \alpha_2 x_2) \leq \alpha_1 f(x_1) + \alpha_2 f(x_2)$$

para $\alpha_1, \alpha_2 \in [0, 1]$, $\alpha_1 + \alpha_2 = 1$.) Es estrictamente convexa si y sólo si en (6.23) se tiene desigualdad estricta cuando $x_1 \neq x_2$ y $t \in (0, 1)$. (Y análogamente para funciones cóncavas cambiando el signo de la desigualdad).

Demostración. La demostración de este resultado está resumida visualmente en la figura 6.8.

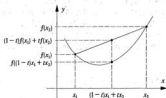


Figura 6.8

- 1.a Supongamos que se cumple (6.23). Sea $\bar{x} \in [x_1, x_2]$. Entonces existe $\bar{t} \in [0, 1]$ tal que $\bar{x} = (1 - \bar{t})x_1 + \bar{t}x_2$ (es el número $\bar{t} = \frac{\bar{x} - x_1}{x_2 - x_1}$, el cual, dado que $x_1 \leq \bar{x} \leq x_2$, verifica que $0 \leq \bar{t} \leq 1$). La recta que pasa por $(x_1, f(x_1))$ y $(x_2, f(x_2))$ tiene por ecuación

$$y - f(x_1) = \frac{f(x_2) - f(x_1)}{x_2 - x_1}(x - x_1)$$

La ordenada del punto de esa recta que tiene por abscisa \bar{x} vale

$$\begin{aligned} \bar{y} &= f(x_1) + (f(x_2) - f(x_1)) \frac{\bar{x} - x_1}{x_2 - x_1} = (\text{según lo anterior}) = \\ &= f(x_1) + (f(x_2) - f(x_1)) \bar{t} = (1 - \bar{t})f(x_1) + \bar{t}f(x_2) \end{aligned}$$

La condición (6.23) expresa entonces que $f(\bar{x})$ es menor o igual que la ordenada del punto de la cuerda que tiene también la abscisa \bar{x} , es decir, que la función es convexa según la definición geométrica dada.

- 1.b Recíprocamente, supongamos que la función es convexa según dicha definición; se tiene que, dado $\bar{t} \in [0, 1]$, el punto

$$\bar{x} = (1 - \bar{t})x_1 + \bar{t}x_2 = x_1 + \bar{t}(x_2 - x_1)$$

pertenece al intervalo $[x_1, x_2]$. Aplicando en él la definición de convexidad se tiene precisamente la desigualdad (6.23).

2. La demostración de las demás afirmaciones se hace de manera análoga.

Puesto que se trata de una condición necesaria y suficiente, podríamos haber adoptado la propiedad (6.23) como definición de función convexa. Así lo haremos efectivamente para el caso general de funciones de varias variables, cuando ya dejamos de disponer de la intuición geométrica del plano y el espacio:

Definición 6.4 Sean $f : \text{Dom } f \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ y K un subconjunto convexo de $\text{Dom } f$. Se dice que f es **convexa en K** si cualesquiera que sean $a, b \in K$ y $t \in [0, 1]$ se tiene que

$$f((1-t)a + tb) \leq (1-t)f(a) + tf(b) \quad (6.24)$$

(equivalentemente

$$f(\alpha_1 a + \alpha_2 b) \leq \alpha_1 f(a) + \alpha_2 f(b)$$

para $\alpha_1, \alpha_2 \in [0, 1]$, $\alpha_1 + \alpha_2 = 1$ y todo par $a, b \in K$). Si en (6.24) se tiene desigualdad estricta cuando $a \neq b$ y $t \in (0, 1)$ se dice que f es **estrictamente convexa**. Las definiciones de función **cóncava** y **estrictamente cóncava** en K se obtienen cambiando en lo anterior el signo de la desigualdad.

La hipótesis incluida en la definición de que K sea un conjunto convexo se necesita para que el punto $(1-t)a + tb$ esté en el dominio de f siempre que lo estén a y b y tenga sentido $f((1-t)a + tb)$. En \mathbb{R} , los conjuntos convexos son los intervalos (véase el apartado 7.7).

Reunimos en el siguiente enunciado varias propiedades de las funciones convexas y cóncavas fáciles de demostrar:

Proposición 6.2

1. f es convexa si y sólo si $-f$ es cóncava.
2. f es convexa si y sólo si cualesquiera que sean $a_1, \dots, a_m \in K$ y $\alpha_1, \dots, \alpha_m \in [0, 1]$, con $\alpha_1 + \dots + \alpha_m = 1$, se verifica que

$$f(\alpha_1 a_1 + \dots + \alpha_m a_m) \leq \alpha_1 f(a_1) + \dots + \alpha_m f(a_m)$$

(con desigualdad estricta para funciones estrictamente convexas y desigualdades opuestas para la concavidad)

3. Si $f_1(x), \dots, f_m(x)$ son funciones convexas (resp. cóncavas) y $\lambda_1, \dots, \lambda_m$ son números ≥ 0 , entonces la función

$$f(x) = \lambda_1 f_1(x) + \dots + \lambda_m f_m(x)$$

es convexa (resp. cóncava).

4. Si f es convexa (resp. cóncava) en K y $c \in \mathbb{R}$, el conjunto

$$K_c = \{x \in K; f(x) \leq c\} \text{ (resp. } \{x \in K; f(x) \geq c\})$$

es convexo. (También son convexos los correspondientes conjuntos con desigualdad estricta)

El recíproco del punto 4 ("Si K_c es convexo para todo $c \in \mathbb{R}$, entonces f es convexa") no es cierto: considérese la función $f(x) = x^3$ en $K = \mathbb{R}$. Este resultado servirá para identificar conjuntos convexos utilizando los criterios de convexidad y concavidad de funciones que veremos enseguida. Así, si las funciones $g_1(x), \dots, g_m(x)$ son convexas en el convexo D , el conjunto

$$\begin{aligned} K &= \{x \in D; g_j(x) \leq c_j, j = 1, \dots, m\} = \\ &= \{x \in D; g_1(x) \leq c_1\} \cap \dots \cap \{x \in D; g_m(x) \leq c_m\} \end{aligned}$$

es convexo por ser intersección de conjuntos convexos.

¿En qué se simplifica la optimización para funciones convexas y cóncavas? Se tiene en primer lugar el siguiente resultado:

Teorema 6.7

1. Si f es convexa (resp. cóncava) en el convexo K , todo mínimo (resp. máximo) local en K es global.
2. Si f es estrictamente convexa (resp. cóncava) en K y alcanza su valor mínimo (resp. máximo) en un punto de K , éste es único.

Demostración. Sea x_0 un mínimo local. Supongamos que existe $x^* \in K$ tal que $f(x^*) < f(x_0)$. Entonces

$$f((1-t)x_0 + tx^*) \leq (1-t)f(x_0) + tf(x^*) < (1-t)f(x_0) + tf(x_0) = f(x_0)$$

y, para t suficientemente pequeño, tendríamos puntos $x = (1-t)x_0 + tx^*$ en cualquier entorno de x_0 en los que $f(x) < f(x_0)$, contradiciendo la definición de mínimo local (figura 6.9).

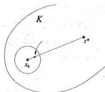


Figura 6.9

En cuanto al punto 2, si la función es estrictamente convexa y alcanzase su valor mínimo m en K en dos puntos distintos x_0 e y_0 se tendría para $t \in (0, 1)$

$$f((1-t)x_0 + ty_0) < (1-t)f(x_0) + tf(y_0) = m$$

es decir, existirían puntos del segmento $[x_0, y_0]$ en los que f valdría menos que m , lo que supone una contradicción.

EJEMPLO 8. La función $f(x_1, x_2) = x_1^2 + x_2^2$ es estrictamente convexa en \mathbb{R}^2 . En efecto, poniendo $a = (a_1, a_2)$, $b = (b_1, b_2)$ se tiene, desarrollando y simplificando,

$$\begin{aligned} & (1-t)f(a) + tf(b) - f((1-t)a + tb) = \\ &= (1-t)(a_1^2 + a_2^2) + t(b_1^2 + b_2^2) - [(1-t)a_1 + tb_1]^2 - [(1-t)a_2 + tb_2]^2 \\ &= t(1-t)[(a_1 - b_1)^2 + (a_2 - b_2)^2] > 0 \text{ si } t \in (0, 1) \text{ y } a \neq b. \end{aligned}$$

$f(x_1, x_2) = x_1^2 + x_2^2$ tiene un mínimo local en $(0, 0)$ que, por lo anterior, será mínimo global (como ya sabíamos, por otra parte, por simples argumentos algebraicos).

Ya se comprende por este ejemplo tan simple que sería muy trabajoso establecer la convexidad o concavidad de una función remitiéndose directamente a la definición. En ésta no se exige que la función sea diferenciable (ni siquiera continua, aunque se puede demostrar que si K es un abierto convexo y f es convexa en K , entonces f es continua en K), pero para funciones diferenciables se tienen las siguientes caracterizaciones; su demostración puede pasarse por alto en una primera lectura:

Teorema 6.8 (condiciones de primer orden) Sea f diferenciable en un conjunto convexo K . Entonces:

1. f es cóncava en K si y sólo si

$$\begin{aligned} f(x) - f(x_0) &\geq df(x_0; x - x_0) = \nabla f(x_0) \cdot (x - x_0) = \\ &= \frac{\partial f}{\partial x_1}(x_0) \cdot (x_1 - x_1^0) + \dots + \frac{\partial f}{\partial x_n}(x_0) \cdot (x_n - x_n^0) \end{aligned} \quad (6.25)$$

para todo par de puntos $x, x_0 \in K$

2. f es estrictamente cóncava si y sólo si

$$\begin{aligned} f(x) - f(x_0) &> df(x_0; x - x_0) = \nabla f(x_0) \cdot (x - x_0) = \\ &= \frac{\partial f}{\partial x_1}(x_0) \cdot (x_1 - x_1^0) + \dots + \frac{\partial f}{\partial x_n}(x_0) \cdot (x_n - x_n^0) \end{aligned} \quad (6.26)$$

para todo par de puntos $x, x_0 \in K$.

(desigualdades opuestas para funciones cóncavas)

Para funciones de una variable, el resultado es intuitivamente obvio (figura 6.10) y viene a significar que la gráfica de una función cóncava está por encima de sus tangentes mientras que la de una función cóncava lo está por debajo.

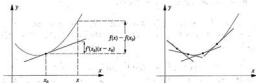


Figura 6.10

Demostración:

1. (a) Sea f cóncava en K y sean $x_0, x \in K$ cualesquiera. Sean $t \in (0, 1)$ y $h = x - x_0$. Entonces

$$f(x_0 + th) = f(x_0 + t(x - x_0)) \leq tf(x_0 + h) + (1 - t)f(x_0)$$

o, equivalentemente,

$$f(x_0 + th) - f(x_0) \leq t(f(x_0 + h) - f(x_0))$$

Restando $df(x_0; th) = t \cdot df(x_0; h)$ en ambos miembros y dividiendo por t

$$\frac{f(x_0 + th) - f(x_0) - df(x_0; th)}{t} \leq f(x_0 + h) - f(x_0) - df(x_0; h)$$

Haciendo $t \rightarrow 0^+$ y teniendo en cuenta la diferenciabilidad de f se obtiene

$$0 \leq f(x_0 + h) - f(x_0) - df(x_0; h)$$

es decir, se verifica (6.25).

- (b) Recíprocamente, supongamos que se satisface (6.25). Sean $a, b \in K$ cualesquiera y pongamos $x_0 = (1-t)a + tb$. Se tiene:

$$\begin{aligned} & (1-t)f(a) + tf(b) - f((1-t)a + tb) = \\ & = (1-t)f(a) + tf(b) - f(x_0) - (1-t)f(x_0) = \\ & = (1-t)(f(a) - f(x_0)) + t(f(b) - f(x_0)) \geq (\text{por (6.25)}) \geq \\ & (1-t)df(x_0; a - x_0) + tdf(x_0; b - x_0) = \\ & = df(x_0; (1-t)(a - x_0) + t(b - x_0)) = \\ & = df(x_0; 0) = 0 \end{aligned}$$

es decir,

$$f((1-t)a + tb) \leq (1-t)f(a) + tf(b) \text{ c.q.d.}$$

2. (a) Supongamos que f es estrictamente convexa. Sean $x, x_0 \in K$, $x \neq x_0$, y $h = x - x_0$. Sea $t \in (0, 1)$. Como f es estrictamente convexa, es convexa y se satisface (6.25). Apliquemos (6.25) con x_0 y $x_0 + th$ en lugar de x . Entonces:

$$\begin{aligned} df(x_0; th) & \leq f(x_0 + th) - f(x_0) = f((1-t)x_0 + tx) - f(x_0) < \\ & < (1-t)f(x_0) + tf(x) - f(x_0) = t(f(x_0 + h) - f(x_0)) \end{aligned}$$

y, dividiendo por t ,

$$df(x_0; x - x_0) < f(x) - f(x_0) \text{ c.q.d.}$$

- (b) El recíproco se demuestra como en el punto 1) sustituyendo el signo " \geq " por " $>$ ".

De este teorema se deduce inmediatamente que para funciones diferenciables convexas (cóncavas) la condición necesaria $\nabla f(x_0) = 0$ de mínimo (máximo) es también suficiente:

Teorema 6.9 Sea f diferenciable y convexa (resp. cóncava) en un conjunto convexo K . Entonces, si $x_0 \in K$ es un punto crítico de f , f alcanza en x_0 su mínimo (resp. máximo) global.

En efecto, se tiene

$$f(x) - f(x_0) \geq \nabla f(x_0) \cdot (x - x_0) = 0 \cdot (x - x_0) = 0$$

es decir, $f(x) \geq f(x_0)$ para todo $x \in K$ (y análogamente para funciones cóncavas).

Corolario 6.2 Si f es estrictamente convexa (o estrictamente cóncava) en K , tiene a lo más un punto crítico en K .

Del teorema 6.8 se deduce también una caracterización muy útil de las funciones convexas y cóncavas de una variable (véase la figura 6.10):

Teorema 6.10 Sea $f: I \rightarrow \mathbb{R}$, donde I es un intervalo de \mathbb{R} . Si existe $f'(x)$ para todo $x \in I$ se tiene:

1. f es convexa en $I \iff f'$ es creciente en I .
2. f es estrictamente convexa en $I \iff f'$ es estrictamente creciente en I .

(para funciones cóncavas se cambia creciente por decreciente).

Demostración.

- a. Supongamos que f es convexa en I y que $x, y \in I$ son tales que $x < y$. Entonces, por (6.25),

$$\begin{aligned} f(y) - f(x) &\geq f'(x)(y - x) \\ f(x) - f(y) &\geq f'(y)(x - y) \Leftrightarrow f(y) - f(x) \leq f'(y)(y - x) \end{aligned}$$

de donde

$$f'(x)(y - x) \leq f'(y)(y - x)$$

y, en consecuencia, por ser $(y - x) > 0$,

$$f'(x) \leq f'(y) \quad (6.27)$$

(Si f es estrictamente convexa resulta $f'(x) < f'(y)$)

- b. Recíprocamente, supongamos que f' es creciente en I , o sea, que se verifica (6.27) siempre que $x < y$. Sean $x_0, x \in I$, $x_0 < x$. Por el teorema del valor medio, existe $\xi \in (x_0, x)$ tal que

$$f(x) - f(x_0) = f'(\xi)(x - x_0)$$

Como f es creciente en I , $f'(\xi) \geq f'(x_0)$, y, por tanto,

$$f(x) - f(x_0) \geq f'(x_0)(x - x_0)$$

es decir, se tiene (6.25) si $x_0 < x$; para $x_0 > x$ se procede de manera análoga; Y si f' es estrictamente creciente se obtiene (6.26).

Para funciones de clase C^2 se tiene un criterio para establecer la convexidad o concavidad de una función que es el que realmente se utiliza en la práctica:

Teorema 6.11 (condiciones de segundo orden) Sea f de clase C^2 en un conjunto abierto y convexo $K \subseteq \mathbb{R}^n$. Entonces:

1. f es convexa en $K \iff d^2f(x; h)$ es semidefinida positiva para todo $x \in K$.
2. Si $d^2f(x; h)$ es definida positiva para todo $x \in K \implies f$ es estrictamente convexa en K .
3. f es cóncava en $K \iff d^2f(x; h)$ es semidefinida negativa para todo $x \in K$.
4. Si $d^2f(x; h)$ es definida negativa para todo $x \in K \implies f$ es estrictamente cóncava en K .

Demostración. Puesto que $f \in C^2(K)$ y K es convexo, se tienen para todo par de puntos $x_0, x \in K$, $h = x - x_0$, las fórmulas de Taylor

$$f(x) - f(x_0) = df(x_0; h) + \frac{1}{2}d^2f(x_0 + \theta h; h) \quad (6.28)$$

$$f(x) - f(x_0) = df(x_0; h) + \frac{1}{2}d^2f(x_0; h) + R_2(h) \quad (6.29)$$

Si $d^2f(y; h) \geq 0$ (resp. > 0) para todo $y \in K$, se tiene, tomando $y = x_0 + \theta h$ en (6.28),

$$f(x) - f(x_0) - df(x_0; x - x_0) \geq 0 \quad (\text{resp. } > 0)$$

es decir, f es convexa (resp. estrictamente convexa) por el teorema 6.8.

Supongamos ahora que f es convexa. Hay que ver que $d^2f(x; h) \geq 0$ para todo $x \in K$ y todo $h \in \mathbb{R}^n$. Supongamos que no es así, o sea, que existen $x_0 \in K$ y $h_0 \in \mathbb{R}^n$, $h_0 \neq 0$, tales que

$$d^2f(x_0; h_0) < 0 \quad (6.30)$$

Como en los criterios de mínimos y máximos locales, de (6.30) se tendría, por (6.29),

$$f(x_0 + th) - f(x_0) - df(x_0; th) < 0$$

para t suficientemente pequeño, lo que implica, por el teorema 6.8, que f no sería convexa, contradiciendo la hipótesis. Las afirmaciones relativas a funciones cóncavas se derivan de las anteriores aplicadas a $-f$. (Que los argumentos sean los mismos que los utilizados para caracterizar mínimos y máximos locales no es de extrañar: tener un mínimo local en x_0 quiere decir que la función es convexa en un ε -entorno de x_0 , y lo análogo para un máximo. Las condiciones 2) y 4) son suficientes pero no necesarias; considérese, por ejemplo, en una dimensión, $f(x) = x^4$.)

Un problema de optimización

$$\underset{x \in K}{\text{opt}} f(x)$$

se dirá que es **convexo** si el conjunto K donde se plantea es convexo y la función $f(x)$ es convexa si se trata de un problema de minimización y cóncava si lo es de maximización. Hemos visto que, según habíamos adelantado, en un problema convexo todo óptimo local es de hecho global y que las condiciones necesarias de primer orden son también suficientes (hay que matizar que esto último no significa, en la práctica, que no haya que recurrir a las derivadas segundas pues lo normal es que la concavidad/convexidad de la función se establezca mediante las condiciones de segundo orden del teorema 6.11).

EJEMPLO 9. En el ejemplo 6 vimos que la función

$$f(x, y) = x^3 + 3xy^2 - 3x^2 - 3y^2 + 4$$

tiene 4 puntos críticos:

$$P_1(0, 0), P_2(2, 0), P_3(1, 1), P_4(1, -1)$$

y que $P_1(0, 0)$ es un máximo local, $P_2(2, 0)$ un mínimo local y $P_3(1, 1)$ y $P_4(1, -1)$ puntos de silla. Nos planteamos hacer un estudio de la concavidad/convexidad de $f(x, y)$ para intentar aplicarlo, junto con lo obtenido allí, a la detección de conjuntos $D \subseteq \mathbb{R}^2$ en los que se pueda afirmar la existencia de óptimos globales de f .

La matriz hessiana de $f(x, y)$ es

$$D^2 f(x, y) = \begin{pmatrix} f''_{xx}(x, y) & f''_{xy}(x, y) \\ f''_{xy}(x, y) & f''_{yy}(x, y) \end{pmatrix} = \begin{pmatrix} 6x - 6 & 6y \\ 6y & 6x - 6 \end{pmatrix}$$

Se tiene

$$H_1(x, y) = 6x - 6, \quad H_2(x, y) = 36(x - 1 + y)(x - 1 - y)$$

Vemos que los conjuntos

$$K_1 = \{(x, y); x > 1, y < x - 1, y > -x + 1\}$$

y

$$K_2 = \{(x, y); x < 1, y > x - 1, y < -x + 1\}$$

son abiertos y convexos (son intersección de semiplanos) y que en K_1 $d^2f(x; h)$ es definida positiva mientras que en K_2 es definida negativa. Por lo tanto, $(2, 0)$ es un mínimo global estricto de f relativamente a K_1 y $(0, 0)$ un máximo global estricto relativamente a K_2 (Figura 6.11).

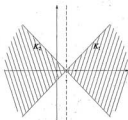


Figura 6.11

Aplicación: el método de mínimos cuadrados

Supongamos que una magnitud y depende de otra magnitud x y que en una serie de experimentos se han obtenido los valores y_1, \dots, y_n correspondientes a los valores x_1, \dots, x_n . Al representarlos en el plano cartesiano, se observa que los puntos $(x_1, y_1), \dots, (x_n, y_n)$ están sensiblemente alineados, lo que lleva a la conjetura de que hay una relación lineal entre las variables x e y (figura 6.12). Se trata, entonces, de determinar la recta $y = mx + b$ —el modelo lineal— que mejor se ajusta a esos datos obtenidos experimentalmente. El criterio para ello es el de que sea mínimo el error entre los valores observados y_i y los que proporciona la fórmula $y = mx + b$ para $x = x_i$. Para cada i , ese error es $|mx_i + b - y_i|$; como trabajar con el valor absoluto tiene sus dificultades,

consideramos en su lugar el cuadrado de ese error: $(mx_i + b - y_i)^2$; la suma de los cuadrados de los errores para n datos observados es

$$E = \sum_{i=1}^n (mx_i + b - y_i)^2$$

cantidad que, dados (x_i, y_i) , nos proponemos minimizar como función (cuadrática) de las variables m y b .

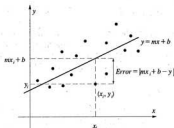


Figura 6.12

Derivando se obtiene

$$\frac{\partial E}{\partial m} = \sum_{i=1}^n 2(mx_i + b - y_i)x_i$$

$$\frac{\partial E}{\partial b} = \sum_{i=1}^n 2(mx_i + b - y_i)$$

Reordenando términos, resulta que los puntos críticos de E son las soluciones del sistema lineal

$$\begin{cases} m \sum_{i=1}^n x_i^2 + b \sum_{i=1}^n x_i = \sum_{i=1}^n x_i y_i \\ m \sum_{i=1}^n x_i + bn = \sum_{i=1}^n y_i \end{cases}$$

El determinante de los coeficientes es

$$n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2$$

cantidad que es ≥ 0 , por la desigualdad de Cauchy (apartado 1.4) aplicada a los vectores $x = (x_1, \dots, x_n)$ y $(1, \dots, 1)$:

$$|((x_1, \dots, x_n), (1, \dots, 1))| = \sum_{i=1}^n |x_i| \leq \sqrt{n} \left(\sum_{i=1}^n x_i^2 \right)^{1/2}$$

De hecho, sólo puede valer cero si todos los valores x_i son iguales (véase el apartado 1.4), lo que no tiene ningún interés en esta aplicación práctica. Así pues, se puede suponer que siempre es $n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2 > 0$ y que, en consecuencia, $E(\bar{m}, \bar{b})$ tiene un único punto crítico (\bar{m}, \bar{b}) .

Por otro lado

$$\frac{\partial^2 E}{\partial m^2} = 2 \sum_{i=1}^n x_i^2, \quad \frac{\partial^2 E}{\partial m \partial b} = 2 \sum_{i=1}^n x_i = \frac{\partial^2 E}{\partial b \partial m}, \quad \frac{\partial^2 E}{\partial b^2} = 2n$$

de donde

$$HE(m, b) = \begin{pmatrix} 2 \sum_{i=1}^n x_i^2 & 2 \sum_{i=1}^n x_i \\ 2 \sum_{i=1}^n x_i & 2n \end{pmatrix}$$

Se tiene:

$$2 \sum_{i=1}^n x_i^2 > 0, \quad 2n > 0, \quad \det(HE(m, b)) = 4 \left(n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2 \right) > 0$$

(por la hipótesis anterior); es decir, $HE(m, b)$ —que no depende de m, b — es definida positiva y, por tanto, $E(m, b)$ es estrictamente convexa en todo \mathbb{R}^2 . Se trata, pues, de un problema convexo y el único punto crítico (\bar{m}, \bar{b}) proporciona un mínimo global (estricto) de la función E . La correspondiente recta $y = \bar{m}x + \bar{b}$ que mejor se ajusta a los datos obtenidos (x_i, y_i) se llama **recta de regresión** y su interés está en que podrá utilizarse para predecir valores de y correspondientes a valores de x que no han sido utilizados o evaluados en el experimento.

6.2 Restricciones de igualdad

Con frecuencia, los problemas de optimización se plantean como una toma de decisión sobre los valores que han de asignarse a las variables que puede

controlar el "optimizador" para que determinada magnitud que depende de dichas variables alcance su valor óptimo (minimización de costes en un proceso productivo, maximización de beneficios, etc.). En muchas aplicaciones, esas **variables de decisión** están obligadas a satisfacer ciertas relaciones funcionales, o sea, no se pueden elegir independientes entre sí y ello limita las posibilidades de decisión del optimizador.

Supongamos, por ejemplo, el caso de un ingeniero de una planta industrial que ha de aprovechar una plancha de acero sobrante para construir un depósito abierto de base cuadrada que almacene la mayor cantidad posible de cierto líquido.

Si x es el lado de la base e y la altura del depósito, habrá que maximizar el volumen $V(x, y) = x^2y$ eligiendo apropiadamente las dimensiones x e y . x e y —variables de decisión— no son en realidad independientes entre sí ya que la superficie $S(x, y) = x^2 + 4xy$ ha de valer una cantidad fija a dada. El problema es, por tanto,

$$\text{maximizar } \{x^2y\}$$

sujeta a la restricción o ligadura

$$x^2 + 4xy = a$$

Este es un ejemplo típico de **problema de optimización con una restricción de igualdad**. Para funciones de dos variables, simbolizamos este tipo de problemas por

$$\begin{aligned} \text{opt } f(x, y); (x, y) \in S \\ \text{s.a. } g(x, y) = 0 \end{aligned} \quad (6.31)$$

(leemos: optimizar —maximizar o minimizar, según de lo que se trate— la función $f(x, y)$ en S sujeta a la restricción de igualdad $g(x, y) = 0$).

La restricción tiene el efecto de reducir el dominio (y, por tanto, también el recorrido) de la **función objetivo** $f(x, y)$ (o sea, la función que se trata de optimizar): se reducen los **grados de libertad** del problema. El conjunto

$$B = \{(x, y) \in S; g(x, y) = 0\}$$

se denomina **conjunto de soluciones factibles** (o, simplemente, **conjunto factible**) del problema de optimización. En general, los puntos que satisfacen $g(x, y) = 0$ forman, como sabemos, una curva en el plano. En el ejemplo anterior, S , el conjunto de \mathbb{R}^2 en donde se plantea en principio el problema de optimización, es, de manera natural, el primer cuadrante abierto $\{(x, y) \in \mathbb{R}^2; x > 0, y > 0\}$, por la naturaleza geométrica de las variables x e

y. Al añadir la restricción sólo cuentan los valores de (x, y) que la satisfacen, o sea, los puntos de la curva $x^2 + 4xy = a$ que están en el primer cuadrante (Fig. 6.13) y los valores de $V(x, y) = x^2y$ que entran en juego para la determinación del máximo de ellos serán los que corresponden a puntos (x, y) que están en esa curva con $0 < x < \sqrt{a}$.

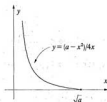


Figura 6.13

En general, para un problema de la forma

$$\begin{aligned} \max f(x, y), \quad (x, y) \in S \\ \text{s.a. } g(x, y) = 0 \end{aligned}$$

se tendrá gráficamente la situación que ilustra la figura 6.14: el máximo restringido será menor que el máximo libre, sin restricciones (y lo contrario para mínimos).

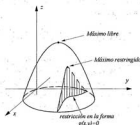


Figura 6.14

Una primera vía de ataque para abordar el problema (6.31) sería utilizar la restricción $g(x, y) = 0$ para “despejar” una de las variables en función de la otra: $y = \varphi(x)$, por ejemplo, y sustituirla en la función objetivo:

$$f(x, y) = f(x, \varphi(x)) = \phi(x)$$

quedando entonces un problema de optimización en una variable (había dos grados de libertad y la ligadura entre x e y los reduce en una unidad). Resolvamos de este modo el ejemplo anterior

$$\begin{aligned} \max \{x^2y\}, \quad (x, y) \in \{(x, y) \in \mathbb{R}^2; x > 0, y > 0\} \\ \text{s.a. } x^2 + 4xy - a = 0 \end{aligned}$$

De $x^2 + 4xy - a = 0$ se obtiene

$$y = \varphi(x) = \frac{a - x^2}{4x} \quad (6.32)$$

con $x \in (0, \sqrt{a})$ pues ha de ser $x > 0, y > 0$; sustituyendo en $f(x, y) = x^2y$ resulta

$$\phi(x) = x^2 \frac{a - x^2}{4x} = \frac{1}{4}x(a - x^2) \quad (6.33)$$

función que hay que maximizar en el intervalo $(0, \sqrt{a})$. Se tiene

$$\phi'(x) = \frac{1}{4}(a - 3x^2)$$

Su único cero en $(0, \sqrt{a})$ es $x^* = \sqrt{a/3}$. Como $\phi''(x) = -\frac{3x}{2} < 0$ para todo $x \in (0, \sqrt{a})$, $\phi(x)$ es estrictamente cóncava en $(0, \sqrt{a})$ y, por tanto, tiene en $x^* = \sqrt{a/3}$ un máximo global. Esto significa que $f(x, y) = x^2y$ alcanza en $(x^*, y^*) = (x^*, \varphi(x^*)) = (\sqrt{a/3}, (1/2)\sqrt{a/3})$ su máximo global en $S = \{(x, y) \in \mathbb{R}^2; x > 0, y > 0\}$ restringido (o condicionado) por la ecuación $x^2 + 4xy = a$, que vale

$$f(x^*, y^*) = \frac{a}{3} \cdot \frac{1}{2} \sqrt{\frac{a}{3}} = \frac{a}{6} \sqrt{\frac{a}{3}}$$

Este método de utilizar la restricción para rebajar la dimensión del problema puede resultar útil en casos simples como éste, pero, si bien la ecuación $g(x, y) = 0$ define una variable como función de la otra bajo hipótesis muy generales, como sabemos, puede ocurrir que tal función sea muy engorrosa de manejar o, simplemente, imposible de obtener explícitamente, como, por ejemplo, para la restricción

$$x - \ln x + y - 2 \ln y - 2 = 0$$

Y la complicación aumentaría al intentar resolver por este camino problemas en varias variables y con varias restricciones.

Hay un método alternativo, el método de los multiplicadores de Lagrange, que, generalmente, lleva a cálculos más sencillos y que además proporciona en ciertas aplicaciones (como veremos más adelante) información

complementaria muy valiosa sobre el problema de optimización bajo estudio. Veamos: se trata de obtener condiciones **necesarias** de óptimo análogas a las de primer orden de los problemas sin restricciones. Supongamos que S es abierto y que f y g son de clase C^1 en S . Supongamos que se alcanza un óptimo (máximo o mínimo) en $(x^*, y^*) \in B = \{(x, y) \in S; g(x, y) = 0\}$. Si

$$\frac{\partial g}{\partial y}(x^*, y^*) \neq 0 \quad (6.34)$$

entonces $g(x, y) = 0$ define implícitamente, por el teorema de la función implícita, una única función $y = \varphi(x)$ con $y^* = \varphi(x^*)$ en un entorno de x^* y además

$$\frac{d\varphi}{dx}(x^*) = -\frac{\frac{\partial g}{\partial x}(x^*, y^*)}{\frac{\partial g}{\partial y}(x^*, y^*)} \quad (6.35)$$

Por otro lado, si, como se supone, la función

$$x \mapsto \phi(x) = f(x, \varphi(x))$$

tiene un óptimo en x^* , habrá de ser (condición necesaria de primer orden)

$$\frac{d\phi}{dx}(x^*) = \frac{\partial f}{\partial x} + \frac{\partial f}{\partial y} \frac{d\varphi}{dx} \Big|_{(x^*, y^*)} = 0 \quad (6.36)$$

o sea,

$$\frac{\partial f}{\partial x}(x^*, y^*) - \frac{\frac{\partial f}{\partial y}(x^*, y^*)}{\frac{\partial g}{\partial y}(x^*, y^*)} \cdot \frac{\partial g}{\partial x}(x^*, y^*) = 0 \quad (6.37)$$

La ecuación (6.37) junto con la ecuación $g(x, y) = 0$ podrían, en principio, resolverse en las variables x, y para dar puntos (x^*, y^*) *candidatos* (condición necesaria) a proporcionar soluciones del problema (6.31) (obsérvese que (6.37) es una condición que ha de satisfacer todo máximo o mínimo local de f en B).

Reordenemos esto en una forma más útil en la práctica, sobre todo cuando abordemos problemas en varias variables con más de una restricción. Definimos el número λ^* por

$$\lambda^* = \frac{\frac{\partial f}{\partial y}(x^*, y^*)}{\frac{\partial g}{\partial y}(x^*, y^*)} \quad (6.38)$$

Entonces, en el óptimo (x^*, y^*) ha de verificarse

$$\begin{cases} \frac{\partial f}{\partial x} - \lambda \frac{\partial g}{\partial x} = 0 \\ \frac{\partial f}{\partial y} - \lambda \frac{\partial g}{\partial y} = 0 \\ g(x, y) = 0 \end{cases} \quad (6.39)$$

para un cierto valor λ^* de λ . Equivalentemente

$$\begin{cases} \nabla f(x^*, y^*) = \lambda^* \nabla g(x^*, y^*) \\ g(x^*, y^*) = 0 \end{cases} \quad (6.40)$$

La condición necesaria (6.40) tiene un significado geométrico muy claro: en un óptimo de f restringido por $g(x, y) = 0$, la curva de nivel $f(x, y) = f(x^*, y^*)$ que pasa por él (si está definida) es tangente a la curva $g(x, y) = 0$. (Figura 6.15; recuérdese que el gradiente de una función es perpendicular a sus curvas de nivel).

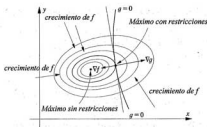


Figura 6.15

Es interesante profundizar en la necesidad de la condición (6.40) desde un punto de vista geométrico: si en un máximo (por ejemplo), las curvas $f(x, y) = f(x^*, y^*) = C^*$ y $g(x, y) = 0$ se cortasen transversalmente (figura 6.16), al ir recorriendo la curva $g = 0$ en el sentido indicado en la figura 6.16 iríamos cortando curvas de nivel correspondientes a valores cada vez mayores de la función objetivo $f(x, y)$, sobrepasaríamos el punto (x^*, y^*) y obtendríamos puntos de dicha curva (el conjunto factible) en los que la función f valdría más que en (x^*, y^*) , lo que va contra la hipótesis de que en (x^*, y^*) se alcanza un máximo.

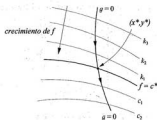


Figura 6.16

λ puede ser ≥ 0 ó ≤ 0 , es decir, $\nabla f(x^*, y^*)$ es múltiplo de $\nabla g(x^*, y^*)$ pero puede tener el mismo sentido o sentido opuesto.

Para generar y aplicar las condiciones necesarias (6.39) es útil introducir la función auxiliar de tres variables

$$L(x, y, \lambda) = f(x, y) - \lambda g(x, y) \quad (6.41)$$

$L(x, y, \lambda)$ se denomina **función lagrangiana** (o, simplemente, **lagrangiano**) del problema. La variable auxiliar λ es el **multiplicador de Lagrange** asociado a la restricción $g(x, y) = 0$. Obsérvese que $L \equiv f$ en los puntos del conjunto factible, o sea, los que satisfacen $g(x, y) = 0$.

Las condiciones (6.39) se pueden escribir en la forma

$$\begin{cases} \frac{\partial L}{\partial x}(x^*, y^*, \lambda^*) = \frac{\partial f}{\partial x}(x^*, y^*, \lambda^*) - \lambda^* \frac{\partial g}{\partial x}(x^*, y^*, \lambda^*) = 0 \\ \frac{\partial L}{\partial y}(x^*, y^*, \lambda^*) = \frac{\partial f}{\partial y}(x^*, y^*, \lambda^*) - \lambda^* \frac{\partial g}{\partial y}(x^*, y^*, \lambda^*) = 0 \\ \frac{\partial L}{\partial \lambda}(x^*, y^*, \lambda^*) = -g(x^*, y^*) = 0 \end{cases} \quad (6.42)$$

que muestra que (x^*, y^*, λ^*) ha de ser un punto crítico del lagrangiano del problema. Se sigue entonces el siguiente esquema de trabajo para intentar resolver el problema (6.31):

1. Se forma el lagrangiano del problema

$$L(x, y, \lambda) = f(x, y) - \lambda g(x, y)$$

2. Se resuelve el sistema de puntos críticos del lagrangiano

$$\begin{cases} \frac{\partial L}{\partial x} = \frac{\partial f}{\partial x} - \lambda \frac{\partial g}{\partial x} = 0 \\ \frac{\partial L}{\partial y} = \frac{\partial f}{\partial y} - \lambda \frac{\partial g}{\partial y} = 0 \\ \frac{\partial L}{\partial \lambda} = -g(x, y) = 0 \end{cases} \quad (6.43)$$

Sus soluciones (x^*, y^*, λ^*) —llamadas *puntos estacionarios*— proporcionan, quedándose con las dos primeras coordenadas (x^*, y^*) , puntos de *B candidatos* a soluciones (óptimos) del problema (6.31).

3. Si el problema (6.31) tiene solución, ha de estar entre esos candidatos: se evalúa la función f en ellos y, si el problema es uno de maximización el mayor de los valores obtenidos da la solución y, si es de minimización, la da el menor de ellos. Para dilucidar si efectivamente existe solución habrá que apelar a argumentos suplementarios, generalmente basados en el teorema de Weierstrass o en propiedades de convexidad que enseguida comentaremos. Por otra parte, téngase presente que en la deducción anterior suponíamos que S , el conjunto de partida en la optimización, era *abierto*. Si no lo es, habrá que considerar también la evaluación de f en los posibles puntos de la frontera de S que satisfacen $g(x, y) = 0$, pues lo que hemos demostrado es que la condición (6.40) se ha de verificar si el óptimo (x^*, y^*) es un *punto interior* de S .

EJEMPLO 10. (Compárese con el ejemplo 2)

$$\begin{cases} \text{opt } \{xy\}, (x, y) \in \mathbb{R}^2 \\ \text{s.a. } x^2 + y^2 = 1 \end{cases}$$

El lagrangiano es

$$L(x, y, \lambda) = xy - \lambda(x^2 + y^2 - 1)$$

Los puntos críticos del lagrangiano son las soluciones del sistema

$$\begin{cases} \frac{\partial L}{\partial x} = y - 2\lambda x = 0 \\ \frac{\partial L}{\partial y} = x - 2\lambda y = 0 \\ \frac{\partial L}{\partial \lambda} = -(x^2 + y^2 - 1) = 0 \end{cases}$$

De las dos primeras ecuaciones se deduce $x^2 = y^2$ (ha de ser $\lambda \neq 0$, pues, si no, sería $x = y = 0$, que no satisface la tercera, y asimismo $x \neq 0, y \neq 0$). Sustituyendo en la tercera se obtienen cuatro puntos críticos del lagrangiano:

$$\left(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}, \frac{1}{2}\right); \left(-\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}, -\frac{1}{2}\right); \left(-\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2}, \frac{1}{2}\right); \left(\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2}, -\frac{1}{2}\right)$$

que proporcionan cuatro candidatos a solución:

$$A\left(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\right); B\left(-\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\right); C\left(-\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2}\right); D\left(\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2}\right)$$

El conjunto factible es la circunferencia $x^2 + y^2 = 1$, que es un conjunto cerrado y acotado. Por tanto, hay solución del problema por el teorema de Weierstrass. Vemos que

$$f(A) = \frac{1}{2}; f(B) = -\frac{1}{2}; f(C) = \frac{1}{2}; f(D) = -\frac{1}{2}$$

con lo que el máximo se alcanza en A y C y el mínimo en B y D (figura 6.4).

EJEMPLO 11. (sobre la condición de regularidad)

$$\begin{cases} \min\{x^2 + y^2\}, (x, y) \in \mathbb{R}^2 \\ \text{s.a. } (x+1)^3 + y^2 = 0 \end{cases} \quad (6.44)$$

El lagrangiano es

$$L(x, y, \lambda) = x^2 + y^2 - \lambda[(x+1)^3 + y^2]$$

y sus puntos críticos satisfacen

$$\begin{aligned} \frac{\partial L}{\partial x} &= 2x - 3\lambda(x+1)^2 = 0 \\ \frac{\partial L}{\partial y} &= 2y - 2\lambda y = 0 \\ \frac{\partial L}{\partial \lambda} &= -[(x+1)^3 + y^2] = 0 \end{aligned} \quad (6.45)$$

Factorizando en la segunda ecuación: $2y(1 - \lambda) = 0$; con $y = 0$, se obtiene en la tercera ecuación $x = -1$, pero estos valores no satisfacen la primera; por su parte, sustituyendo $\lambda = 1$ en la primera ecuación resulta $3x^2 + 4x + 3 = 0$, cuyas raíces son complejas. Así pues, el sistema (6.45) en las variables (x, y, λ) no tiene soluciones, con lo que, aparentemente, no hay candidatos a solución

del problema de minimización propuesto. Sin embargo, es fácil ver gráficamente que el problema (6.44) tiene solución (Fig. 6.17): $(-1, 0)$, el punto más próximo al origen de la curva $(x+1)^3 + y^2 = 0$. La explicación es que en él no se satisface la **condición de regularidad**

$$\frac{\partial g}{\partial y}(x^*, y^*) \neq 0$$

que tomábamos como hipótesis (con objeto de aplicar el teorema de la función implícita) para obtener la *condición necesaria* (6.40). Hay, por tanto, que estudiar aparte los puntos en que no se cumple esa hipótesis de regularidad y tenerlos en cuenta como posibles soluciones del problema de optimización.

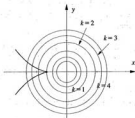


Figura 6.17

EJEMPLO 12. (óptimos locales y globales)

$$\begin{aligned} &\text{opt } \{2x^3 + 9y^2 + 12x\}, \quad (x, y) \in \mathbb{R}^2 \\ &\text{s.a. } x + y = 0 \end{aligned} \quad (6.46)$$

El lagrangiano es

$$L(x, y, \lambda) = 2x^3 + 9y^2 + 12x - \lambda(x + y)$$

y su sistema de puntos críticos

$$\begin{aligned} 6x^2 + 12 - \lambda &= 0 \\ 18y - \lambda &= 0 \\ x + y &= 0 \end{aligned}$$

Resolviéndolo se obtienen dos soluciones:

$$(-1, 1, 18) \text{ y } (-2, 2, 36)$$

y, por tanto, dos candidatos a solución del problema (6.46): $(-1, 1)$ y $(-2, 2)$ (en este caso, la condición de regularidad se satisface en todo punto del conjunto factible B). Se tiene $f(-1, 1) = -5$ y $f(-2, 2) = -4$, con lo que, aparentemente, $(-1, 1)$ proporciona el mínimo y $(-2, 2)$ el máximo. Pero, en realidad, *no hay solución* del problema: utilizando la restricción para sustituir $y = -x$ en $f(x, y)$ se tiene $f(x, -x) = 2x^3 + 9x^2 + 12x$ y vemos que

$$\lim_{x \rightarrow \infty} f(x, -x) = \infty \text{ y } \lim_{x \rightarrow -\infty} f(x, -x) = -\infty$$

Estudiando $\phi(x) = 2x^3 + 9x^2 + 12x$, $x \in \mathbb{R}$, se observa que $x_1 = -1$ y $x_2 = -2$ son sus puntos críticos y que $\phi''(-1) = 6 > 0$ y $\phi''(-2) = -6 < 0$, lo que significa que $(-1, 1)$ es un mínimo local y $(-2, 2)$ un máximo local de $f(x, y) = 2x^3 + 9y^2 + 12x$ sujeta a la restricción $x + y = 0$ (como señalamos, la condición necesaria (6.40) ha de ser satisfecha por todo óptimo local en el que se verifique la condición de regularidad (6.34)). (EJERCICIO: En relación a las cuestiones planteadas en este ejemplo, en particular la apariencia de que el menor valor de f proporciona el mínimo y el mayor valor el máximo, analizar el problema

$$\left\{ \begin{array}{l} \text{opt } \{x(y^2 + 1)\}, \quad (x, y) \in \mathbb{R}^2 \\ \text{s.a. } x^2 - 1 = 0 \end{array} \right\}$$

EJEMPLO 13. (óptimo en la frontera de S)

$$\begin{array}{ll} \max \{3x + 2y\}, & (x, y) \in S = \{(x, y) \in \mathbb{R}^2; x \geq 0, y \geq 0\} \\ \text{s.a. } xy + x + y = 5 \end{array} \quad (6.47)$$

Lagrangiano:

$$L(x, y, \lambda) = 3x + 2y - \lambda(xy + x + y - 5)$$

Puntos críticos del lagrangiano:

$$\begin{array}{rcl} 3 - \lambda(y + 1) & = & 0 \\ 2 - \lambda(x + 1) & = & 0 \\ xy + x + y & = & 5 \end{array}$$

De las dos primeras ecuaciones se obtiene

$$3(x + 1) = 2(y + 1)$$

Sustituyendo en la tercera resulta

$$(x + 1)^2 = 4$$

Como sólo nos interesan soluciones positivas, $x = 1$, de la que se deduce $y = 2$. Así pues, se obtiene como único candidato a solución el punto $(1, 2)$. Se tiene $f(1, 2) = 7$. Pero obsérvese que el punto $(0, 5)$ también está en S , satisface la restricción y $f(0, 5) = 10$, luego $(1, 2)$ no proporciona el máximo. ¿Hay solución del problema (6.47)? Sí, porque el conjunto factible B , puntos de la curva $xy + x + y = 5$ que están en el primer cuadrante cerrado, es cerrado y acotado (Fig. 6.18), y en consecuencia, por el teorema de Weirstrass, la función continua $f(x, y) = 3x + 2y$ alcanza en él su valor máximo y su valor mínimo. Considerando las curvas (rectas) de nivel de $f(x, y) = 3x + 2y$ se ve fácilmente (figura 6.18) que f alcanza en $(1, 2)$ su valor mínimo en B mientras que el máximo lo alcanza en $(5, 0)$ (y vale $f(5, 0) = 15$), punto que está en la frontera de S y no tenía por qué aparecer entre los puntos críticos del lagrangiano.

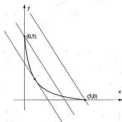


Figura 6.18

El problema general de optimización con restricciones de igualdad es

$$\begin{aligned} &\text{opt } f(x_1, \dots, x_n), \quad (x_1, \dots, x_n) \in S \subseteq \text{Dom } f \\ &\text{s.a.} \\ &\begin{cases} g_1(x_1, \dots, x_n) = 0 \\ g_2(x_1, \dots, x_n) = 0 \\ \vdots \\ g_m(x_1, \dots, x_n) = 0 \end{cases} \end{aligned} \quad (6.48)$$

El conjunto $B = \{(x_1, \dots, x_n) \in S; g_j(x_1, \dots, x_n) = 0, j = 1, \dots, m\}$ es el conjunto de **soluciones factibles** (o **conjunto factible**) del problema de optimización. Obsérvese que si en el caso simple de una función de dos variables

y una restricción que hemos considerado hasta aquí hubiésemos añadido otra restricción que cortase a la primera en uno o varios puntos (Fig. 6.19), el efecto conjunto de ambas sería reducir a esos puntos el conjunto donde optimizar y la determinación del valor óptimo sería un problema algebraico más o menos complicado (dependiendo de la dificultad en resolver el sistema de ecuaciones

$$\begin{aligned}g_1(x, y) &= 0 \\g_2(x, y) &= 0\end{aligned}$$

que puede no tener soluciones) pero no un genuino problema de optimización con restricciones de igualdad.

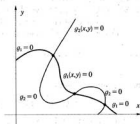


Figura 6.19

Del mismo modo, para que en el caso general se tenga efectivamente uno de tales problemas, el número m de restricciones ha de ser menor que el número n de variables de que depende la función objeto de optimización. Porque, supongamos que $m = n$ y que $x_0 = (x_1^0, \dots, x_n^0)$ es una solución de (6.48); si el jacobiano $\partial(g_1, \dots, g_m)/\partial(x_1, \dots, x_n)$ es $\neq 0$ en x_0 , entonces, por el teorema de la función inversa, en un entorno de x_0 no hay más soluciones de (6.48), o sea, x_0 sería una solución factible aislada como las de la figura 6.19. Y si $\partial(g_1, \dots, g_m)/\partial(x_1, \dots, x_n)$ es nulo en un entorno de x_0 , algunas de las funciones g_1, \dots, g_m son dependientes de las demás en un entorno de x_0 y, en consecuencia, algunas de las ecuaciones de (6.48) serían o bien redundantes —y entonces las restricciones en este entorno serían, en realidad, en número menor que n — o bien incompatibles —y entonces el conjunto factible sería vacío— con las restantes. Finalmente, si $m > n$, por lo menos $m - n$ de las funciones g_1, \dots, g_m serían funcionalmente dependientes de las restantes, al menos en un subconjunto de S , como vimos en el capítulo 4, y estaríamos en la misma situación que para $m = n$. Resulta, pues, razonable, plantear en general este tipo de problemas con la condición $m < n$.

En el capítulo 9 demostraremos el siguiente

Teorema 6.12 (Lagrange) Sean $f, g_1, \dots, g_m, m < n$, funciones de clase C^1 en un subconjunto abierto $S \subseteq \mathbb{R}^n$ con valores en \mathbb{R} . Supongamos que $x^* = (x_1^*, \dots, x_n^*)$ es un óptimo local de f en el conjunto

$$B = \{(x_1, \dots, x_n) \in S; g_j(x_1, \dots, x_n) = 0, j = 1, \dots, m\}$$

y supongamos también que los vectores $\nabla g_1(x^*), \dots, \nabla g_m(x^*)$ son linealmente independientes. Entonces existen constantes $\lambda_1^*, \dots, \lambda_m^*$ tales que

$$\nabla f(x^*) = \lambda_1^* \nabla g_1(x^*) + \dots + \lambda_m^* \nabla g_m(x^*) \quad (6.49)$$

La condición de regularidad en el caso general es la independencia lineal de los vectores $\nabla g_1(x^*), \dots, \nabla g_m(x^*)$ (que nos permitirá, como en el caso simple anterior, aplicar el teorema de la función implícita en su formulación general). Hay otras condiciones de regularidad algo menos exigentes en las que no entraremos aquí.

Para utilizar este teorema en la práctica se siguen los mismos pasos que antes:

1. Se forma el lagrangiano del problema:

$$L(x, \lambda) = f(x_1, \dots, x_n) - \lambda_1 g_1(x_1, \dots, x_n) - \dots - \lambda_m g_m(x_1, \dots, x_n)$$

($\lambda = (\lambda_1, \dots, \lambda_m)$; λ_j es el multiplicador de Lagrange asociado a la restricción $g_j(x) = 0$).

2. Se resuelve el sistema de puntos críticos del lagrangiano

$$\left\{ \begin{array}{l} \frac{\partial L}{\partial x_1} = \frac{\partial f}{\partial x_1} - \sum_{j=1}^m \lambda_j \frac{\partial g_j}{\partial x_1} = 0 \\ \vdots \\ \frac{\partial L}{\partial x_n} = \frac{\partial f}{\partial x_n} - \sum_{j=1}^m \lambda_j \frac{\partial g_j}{\partial x_n} = 0 \\ \frac{\partial L}{\partial \lambda_1} = -g_1(x) = 0 \\ \vdots \\ \frac{\partial L}{\partial \lambda_m} = -g_m(x) = 0 \end{array} \right. \quad (6.50)$$

Como vemos, es un sistema de $n+m$ ecuaciones en $n+m$ incógnitas. Sus soluciones proporcionan, quedándose con las n primeras coordenadas: $x^* = (x_1^*, \dots, x_n^*)$, candidatos a soluciones del problema (6.48).

3. Se procede como en el punto 3 anterior, con las observaciones allí efectuadas.

Problemas convexos

El problema

$$\begin{cases} \min f(x), & x \in S \\ \text{s.a. } g_1(x) = 0, \dots, g_m(x) = 0 \end{cases} \quad (6.51)$$

o sea, el problema de minimizar $f(x)$ en $B = \{x \in S; g_j(x) = 0\}$, es convexo si B es convexo y f es convexa en B . (En la práctica, será frecuente que f sea de clase C^2 en el abierto convexo S y que comprobemos su convexidad por las condiciones de segundo orden del teorema 6.11, y si f es convexa en S , también lo será en B .)

Por el teorema 6.8, se tendrá, bajo esas hipótesis, que

$$f(x+h) - f(x) \geq \nabla f(x) \cdot h \quad (6.52)$$

para todo punto $x \in B$ y todo $h \in \mathbb{R}^n$ tal que $x+h \in B$. Por otro lado, para tales x, h se verifica que

$$\nabla g_j(x) \cdot h = 0 \quad (6.53)$$

En efecto, poniendo $u = h/|h|$ se tiene

$$\begin{aligned} \nabla g_j(x) \cdot h &= |h| (\nabla g_j(x) \cdot u) = |h| D_u g_j(x) = |h| \lim_{s \rightarrow 0} \frac{g_j(x+su) - g_j(x)}{s} = \\ &= (s = t|h|) = \lim_{t \rightarrow 0} \frac{g_j(x+th) - g_j(x)}{t} \end{aligned}$$

Pero si $x, x+h \in B$, entonces $x+th = (1-t)x + t(x+h) \in B$ para todo $t \in [0, 1]$, por ser B convexo. En consecuencia, $g_j(x+th) = g_j(x) = 0$, por la definición de B , y ese límite vale cero. Si suponemos que (x^*, λ^*) es una solución del sistema (6.50), o sea, tal que $g_j(x^*) = 0, j = 1, \dots, m$, y

$$\nabla f(x^*) = \lambda_1^* \nabla g_1(x^*) + \dots + \lambda_m^* \nabla g_m(x^*)$$

para unos ciertos $\lambda_1^*, \dots, \lambda_m^*$, entonces, siendo $h \in \mathbb{R}^n$ tal que $x^*+h \in B$, se tendrá por (6.52) y (6.53)

$$\begin{aligned} f(x^*+h) - f(x^*) &\geq \nabla f(x^*) \cdot h = \\ &= \lambda_1^* (\nabla g_1(x^*) \cdot h) + \dots + \lambda_m^* (\nabla g_m(x^*) \cdot h) = 0 \end{aligned}$$

es decir

$$f(x^*+h) \geq f(x^*)$$

para todo $h \in \mathbb{R}^n$ tal que $x^* + h \in B$, lo que quiere decir que x^* es un mínimo global de f sobre B . Vemos, pues, que si el problema (6.51) es convexo, la condición necesaria de optimalidad local es condición necesaria y suficiente de optimalidad global.

Un caso interesante en el que el conjunto B es convexo es aquél en que las funciones $g_j(x)$ que definen las restricciones son lineales (¿por qué?). Es fácil ver que si f es estrictamente convexa en B , entonces el mínimo global es estricto (y, por tanto, único). Si se trata de un problema de maximización, será convexo si f es cóncava en B , en cuyo caso los máximos serán globales (complétense los detalles).

EJEMPLO 14. Se pide descomponer el número 9 en suma de tres sumandos de manera que la suma de los cuadrados de éstos sea mínima. El problema es

$$\begin{cases} \min \{x^2 + y^2 + z^2\} \\ \text{s.a. } x + y + z = 9 \end{cases}$$

Lagrangiano:

$$L(x, y, z, \lambda) = x^2 + y^2 + z^2 - \lambda(x + y + z - 9)$$

Puntos críticos del lagrangiano:

$$\begin{cases} \frac{\partial L}{\partial x} = 2x - \lambda = 0 \\ \frac{\partial L}{\partial y} = 2y - \lambda = 0 \\ \frac{\partial L}{\partial z} = 2z - \lambda = 0 \\ \frac{\partial L}{\partial \lambda} = -(x + y + z - 9) = 0 \end{cases}$$

La única solución es $(3, 3, 3, 6)$, que proporciona $(3, 3, 3)$ como candidato a solución. Y es la solución efectiva por tratarse de un problema convexo (compruébense los detalles).

Condiciones suficientes de optimalidad

Se pueden establecer para los óptimos con restricciones de igualdad unas condiciones de segundo orden que permiten discriminar entre máximos y mínimos análogas a las que hemos visto en el apartado 6.1 para los problemas sin restricciones.

Consideramos el caso más simple de una función de dos variables sujeta a una restricción:

$$\begin{cases} \text{opt } f(x, y), (x, y) \in S \\ \text{s.a. } g(x, y) = 0 \end{cases}$$

Suponemos que f y g son de clase $C^2(S)$ y retomamos los argumentos con que introducíamos el método de los multiplicadores de Lagrange: se supone que f alcanza un óptimo en $(x^*, y^*) \in B = \{(x, y) \in S; g(x, y) = 0\}$ y que $g'_y(x^*, y^*) \neq 0$. Entonces $g(x, y) = 0$ define implícitamente una única función $y = \varphi(x)$ con $y^* = \varphi(x^*)$ en un entorno de x^* y la función de una variable

$$x \mapsto \phi(x) = f(x, \varphi(x))$$

tiene un óptimo en x^* . En él ha de verificarse

$$\left. \frac{d\phi}{dx}(x^*) = \frac{\partial f}{\partial x} + \frac{\partial f}{\partial y} \frac{d\varphi}{dx} \right|_{(x^*, y^*)} = 0 \quad (6.54)$$

donde, como sabemos,

$$\frac{d\varphi}{dx}(x^*) = -\frac{g'_x(x^*, y^*)}{g'_y(x^*, y^*)} \quad (6.55)$$

Para clasificar x^* como punto crítico de $\phi(x)$ —lo que equivale a clasificar (x^*, y^*) como óptimo de f sujeta a la restricción $g(x, y) = 0$ — calculamos la derivada segunda $\phi''(x)$:

$$\phi'' = f''_{xx} + f''_{xy}\varphi' + f''_{yx}\varphi' + f''_{yy}(\varphi')^2 + f'_y\varphi'' \quad (6.56)$$

Obtenemos φ'' a partir de (6.55):

$$\begin{aligned} \varphi'' &= -\frac{(g''_{xx} + g'_{xy}\varphi')g'_y - g'_x(g''_{yx} + g''_{yy}\varphi')}{(g'_y)^2} = \\ &= \text{(utilizando (6.55))} = \\ &= -\frac{g''_{xx}(g'_y)^2 - 2g'_{xy}g'_xg'_y + g''_{yy}(g'_x)^2}{(g'_y)^3} \end{aligned} \quad (6.57)$$

Sustituyendo (6.55) y (6.57) en (6.56) resulta:

$$\begin{aligned} \phi'' &= f''_{xx} + 2f''_{xy}\varphi' + f''_{yy}(\varphi')^2 + f'_y\varphi'' = \\ &= \frac{1}{(g'_y)^2} \left(\left(f''_{xx} - \frac{f''_{xy}g'_x}{g'_y} \right) (g'_y)^2 - 2 \left(f''_{xy} - \frac{f''_{yy}g'_x}{g'_y} \right) g'_xg'_y + \right. \\ &\quad \left. + \left(f''_{yy} - \frac{f''_{xy}g'_x}{g'_y} \right) (g'_x)^2 \right) \end{aligned} \quad (6.58)$$

Al evaluar en x^* , recordemos que

$$\lambda^* = \frac{f'_y(x^*, y^*)}{g'_y(x^*, y^*)}$$

es el multiplicador de Lagrange asociado a (x^*, y^*) que se habrá obtenido al resolver el sistema de puntos críticos de la función lagrangiana

$$L(x, y, \lambda) = f(x, y) - \lambda g(x, y)$$

Entonces

$$\begin{aligned} \phi''(x^*) &= \frac{1}{(g'_y(x^*, y^*))^2} (L''_{xx}(g'_y)^2 - 2L''_{xy}g'_xg'_y + L''_{yy}(g'_x)^2)|_{(x^*, y^*, \lambda^*)} = \\ &= -\frac{1}{(g'_y(x^*, y^*))^2} \begin{vmatrix} 0 & g'_x & g'_y \\ g'_x & L''_{xx} & L''_{xy} \\ g'_y & L''_{xy} & L''_{yy} \end{vmatrix}_{(x^*, y^*, \lambda^*)} \end{aligned} \quad (6.59)$$

El determinante

$$|\bar{H}| = \begin{vmatrix} 0 & g'_x & g'_y \\ g'_x & L''_{xx} & L''_{xy} \\ g'_y & L''_{xy} & L''_{yy} \end{vmatrix} \quad (6.60)$$

se llama *hessiano oriado* (se oriá el hessiano de L respecto a (x, y) con las derivadas primeras de la función que define la restricción) y, como vemos, su signo en (x^*, y^*, λ^*) es opuesto al de $\phi''(x^*)$. Resulta así el el siguiente criterio:

Teorema 6.13 Sean f y g funciones de clase C^2 en un abierto S de \mathbb{R}^2 con valores en \mathbb{R} . Supongamos que (x^*, y^*, λ^*) , con $\nabla g(x^*, y^*) \neq (0, 0)$, es un punto crítico del lagrangiano $L(x, y, \lambda) = f(x, y) - \lambda g(x, y)$ (lo que equivale a decir que $g(x^*, y^*) = 0$ y $\nabla f(x^*, y^*) = \lambda^* \nabla g(x^*, y^*)$). Entonces:

- (i) Si $|\bar{H}(x^*, y^*, \lambda^*)| < 0$, (x^*, y^*) es un mínimo local (estricto) de f restringida a $B = \{(x, y) \in S; g(x, y) = 0\}$.
- (ii) Si $|\bar{H}(x^*, y^*, \lambda^*)| > 0$, (x^*, y^*) es un máximo local (estricto) de f restringida a $B = \{(x, y) \in S; g(x, y) = 0\}$.
- (iii) Si $|\bar{H}(x^*, y^*, \lambda^*)| = 0$, el criterio no decide y (x^*, y^*) puede ser un máximo, un mínimo o ninguna de las dos cosas.

EJEMPLO 15. Veamos como se aplica el criterio al problema del ejemplo 10. Se tiene

$$|\bar{H}(x, y, \lambda)| = \begin{vmatrix} 0 & 2x & 2y \\ 2x & -2\lambda & 1 \\ 2y & 1 & -2\lambda \end{vmatrix}$$

con lo que

$$\bar{H}(\sqrt{2}/2, \sqrt{2}/2, 1/2) = 8 > 0, \quad \bar{H}(-\sqrt{2}/2, \sqrt{2}/2, -1/2) = -8 < 0$$

$$\bar{H}(-\sqrt{2}/2, -\sqrt{2}/2, 1/2) = 8 > 0, \quad \bar{H}(\sqrt{2}/2, -\sqrt{2}/2, -1/2) = -8 < 0$$

y confirmamos que $(\sqrt{2}/2, \sqrt{2}/2)$ y $(-\sqrt{2}/2, -\sqrt{2}/2)$ son máximos y $(-\sqrt{2}/2, \sqrt{2}/2)$ y $(\sqrt{2}/2, -\sqrt{2}/2)$ mínimos. (No ha de caerse en el error de intentar clasificar los óptimos con restricciones mediante el hessiano de la función objetivo f ; como se ve en este ejemplo, los óptimos con restricciones no tienen por que ser puntos críticos de f y no tendría ningún sentido intentar analizar el signo de Δf en un entorno de (x^*, y^*) —en el que, además, sólo hay que tener en cuenta los puntos que satisfacen la restricción— mediante el bloque de términos de segundo orden del desarrollo de Taylor pues no tienen por que anularse los de primer orden.)

En el caso de una función de n variables sujeta a una restricción

$$\begin{cases} \text{opt } f(x_1, \dots, x_n), & (x_1, \dots, x_n) \in S \\ \text{s.a. } g(x_1, \dots, x_n) = 0 \end{cases}$$

se procedería de manera análoga: si f alcanza un óptimo en un punto $x^* \in B = \{x \in S; g(x) = 0\}$ y se supone, por ejemplo, que $g'_{x_n}(x^*) \neq 0$, entonces $g(x) = 0$ define implícitamente una única función $x_n = \varphi(x_1, \dots, x_{n-1})$ en un entorno de $(x_1^*, \dots, x_{n-1}^*)$ y la función de $n-1$ variables

$$(x_1^*, \dots, x_{n-1}^*) \mapsto \phi(x_1^*, \dots, x_{n-1}^*) = f(x_1^*, \dots, x_{n-1}^*, \varphi(x_1^*, \dots, x_{n-1}^*))$$

tiene un óptimo local —sin restricciones— en $(x_1^*, \dots, x_{n-1}^*)$. Para clasificarlo —lo que equivale a clasificar x^* como óptimo de f sujeta a la restricción $g(x) = 0$ — recurrimos al desarrollo de Taylor de segundo orden de ϕ en un entorno de $(x_1^*, \dots, x_{n-1}^*)$ y utilizamos las condiciones de segundo orden que hemos establecido en el apartado anterior. Al hacer los cálculos —obteniendo por derivación implícita las derivadas segundas de ϕ — se comprueba que la clasificación de la matriz hessiana de ϕ en $(x_1^*, \dots, x_{n-1}^*)$ puede hacerse en

términos del *hessiano orlado*

$$|\hat{H}| = \begin{vmatrix} 0 & g_{x_1} & g_{x_2} & \cdots & g_{x_n} \\ g_{x_1} & L_{x_1x_1} & L_{x_1x_2} & \cdots & L_{x_1x_n} \\ g_{x_2} & L_{x_2x_1} & L_{x_2x_2} & \cdots & L_{x_2x_n} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ g_{x_n} & L_{x_nx_1} & L_{x_nx_2} & \cdots & L_{x_nx_n} \end{vmatrix}$$

evaluado en (x^*, λ^*) , donde λ^* es el multiplicador de Lagrange asociado a x^* que resulta al resolver el sistema de puntos críticos del lagrangiano $L(x, \lambda) = f(x) - \lambda g(x)$. Se consideran los sucesivos menores principales dominantes a partir del orden 3:

$$|\hat{H}_2| = \begin{vmatrix} 0 & g_{x_1} & g_{x_2} \\ g_{x_1} & L_{x_1x_1} & L_{x_1x_2} \\ g_{x_2} & L_{x_2x_1} & L_{x_2x_2} \end{vmatrix}, |\hat{H}_3| = \begin{vmatrix} 0 & g_{x_1} & g_{x_2} & g_{x_3} \\ g_{x_1} & L_{x_1x_1} & L_{x_1x_2} & L_{x_1x_3} \\ g_{x_2} & L_{x_2x_1} & L_{x_2x_2} & L_{x_2x_3} \\ g_{x_3} & L_{x_3x_1} & L_{x_3x_2} & L_{x_3x_3} \end{vmatrix}, \dots$$

siendo el último $|\hat{H}_n| = |\hat{H}|$ (los subíndices se refieren al orden del menor principal del hessiano respecto a x_1, \dots, x_n de la función $L(x, \lambda)$ que se orla con las correspondientes derivadas primeras de la función que define la restricción). Se tiene entonces el siguiente criterio:

- (i) Si $|\hat{H}_2| < 0$, $|\hat{H}_3| < 0, \dots, |\hat{H}_n| < 0$, x^* es un mínimo local (estricto) de f restringida a $B = \{x \in S; g(x) = 0\}$.
- (ii) Si $|\hat{H}_2| > 0$, $|\hat{H}_3| < 0$, $|\hat{H}_4| > 0, \dots$, es decir, $|\hat{H}_2|$ es positivo y los siguientes menores van alternando de signo, entonces x^* es un máximo local (estricto) de f restringida a $B = \{x \in S; g(x) = 0\}$.

EJEMPLO 16. Se trata de determinar entre todos los paralelepípedos de superficie dada a el que tiene volumen máximo (por ejemplo: ¿cómo ha de construirse una caja rectangular cerrada con una cantidad de material dada para conseguir que su volumen sea máximo?). El problema es

$$\begin{cases} \max\{xyz\}, (x, y, z) \in \mathbb{R}_{+++}^3 = \{(x, y, z) \in \mathbb{R}^3; x > 0, y > 0, z > 0\} \\ \text{s.a. } 2xy + 2yz + 2zx = a \end{cases}$$

Lagrangiano:

$$L(x, y, z, \lambda) = xyz - \lambda(2xy + 2yz + 2zx - a)$$

Puntos críticos del lagrangiano:

$$\begin{cases} \frac{\partial L}{\partial x} = yz - 2\lambda(y+z) = 0 \\ \frac{\partial L}{\partial y} = xz - 2\lambda(x+z) = 0 \\ \frac{\partial L}{\partial z} = xy - 2\lambda(x+y) = 0 \\ 2xy + 2yz + 2zx = a \end{cases}$$

sistema que, por ser x, y, z distintas de cero, es equivalente a

$$\begin{cases} xyz - 2\lambda(xy + xz) = 0 \\ xyz - 2\lambda(xy + yz) = 0 \\ xyz - 2\lambda(xz + yz) = 0 \\ 2xy + 2yz + 2zx = a \end{cases}$$

Sumando las tres primeras y utilizando la cuarta resulta $3xyz = 2\lambda a$, que implica $\lambda \neq 0$ y, sustituyendo en aquéllas, $x = y = z$; en consecuencia $6x^2 = a$ y $x = y = z = \sqrt{a/6}$. Así, pues, de haber solución ha de ser el cubo de arista $\sqrt{a/6}$. Clasifiquemos este único punto estacionario $(x^*, y^*, z^*) = (\sqrt{a/6}, \sqrt{a/6}, \sqrt{a/6})$, al que corresponde $\lambda^* = \frac{1}{4}\sqrt{a/6}$. El hessiano orlado es

$$|\bar{H}| = \begin{vmatrix} 0 & 2(y+z) & 2(x+z) & 2(x+y) \\ 2(y+z) & 0 & z-2\lambda & y-2\lambda \\ 2(x+z) & z-2\lambda & 0 & x-2\lambda \\ 2(x+y) & y-2\lambda & x-2\lambda & 0 \end{vmatrix}$$

Evaluándolo en $(\sqrt{a/6}, \sqrt{a/6}, \sqrt{a/6}, \frac{1}{24}\sqrt{6a})$:

$$|H| = \begin{vmatrix} 0 & 4\sqrt{a/6} & 4\sqrt{a/6} & 4\sqrt{a/6} \\ 4\sqrt{a/6} & 0 & \frac{1}{2}\sqrt{a/6} & \frac{1}{2}\sqrt{a/6} \\ 4\sqrt{a/6} & \frac{1}{2}\sqrt{a/6} & 0 & \frac{1}{2}\sqrt{a/6} \\ 4\sqrt{a/6} & \frac{1}{2}\sqrt{a/6} & \frac{1}{2}\sqrt{a/6} & 0 \end{vmatrix}$$

Se tiene:

$$|\bar{H}_2| = \begin{vmatrix} 0 & 4\sqrt{a/6} & 4\sqrt{a/6} \\ 4\sqrt{a/6} & 0 & \frac{1}{2}\sqrt{a/6} \\ 4\sqrt{a/6} & \frac{1}{2}\sqrt{a/6} & 0 \end{vmatrix} = \frac{4}{9}\sqrt{6}(\sqrt{a})^3 > 0$$

$$|H_3| = \begin{vmatrix} 0 & 4\sqrt{a/6} & 4\sqrt{a/6} & 4\sqrt{a/6} \\ 4\sqrt{a/6} & 0 & \frac{1}{2}\sqrt{a/6} & \frac{1}{2}\sqrt{a/6} \\ 4\sqrt{a/6} & \frac{1}{2}\sqrt{a/6} & 0 & \frac{1}{2}\sqrt{a/6} \\ 4\sqrt{a/6} & \frac{1}{2}\sqrt{a/6} & \frac{1}{2}\sqrt{a/6} & 0 \end{vmatrix} = -\frac{1}{3}a^2 < 0$$

luego se trata de un máximo. El resultado es coherente con la naturaleza geométrica del problema: vemos que no hay un paralelepípedo de volumen mínimo para una superficie dada; pero, en rigor, sólo hemos probado que $(\sqrt{a/6}, \sqrt{a/6}, \sqrt{a/6})$ es un máximo local, si bien es claro que no podemos construir con una superficie dada finita un paralelepípedo de volumen arbitrariamente grande y que, en consecuencia, $(\sqrt{a/6}, \sqrt{a/6}, \sqrt{a/6})$ proporciona el máximo global $f(\sqrt{a/6}, \sqrt{a/6}, \sqrt{a/6}) = (\sqrt{a/6})^3$. Para justificarlo analíticamente con todo rigor, podemos argumentar como sigue: despejando z en la restricción

$$z = \frac{a - 2xy}{2(y + x)}$$

y sustituyendo en $f(x, y, z) = xyz$ se obtiene la función

$$\phi(x, y) = xy \frac{a - 2xy}{2(y + x)}$$

ϕ está definida en el primer cuadrante abierto del plano, si bien en nuestro problema interesa su restricción a los puntos que están por debajo de la rama de la hipérbola $xy = a/2$, pues ha de ser $z > 0$. Por otra parte (con la idea de "controlar" los valores de ϕ para valores o muy pequeños o muy grandes de x e y) se tiene que para los puntos (x, y) tales que

$$\frac{xy}{x+y} \leq \frac{M}{a} \text{ donde } M = f_{\max} = (\sqrt{a/6})^3$$

o sea, que están en y por debajo de la rama de la hipérbola $y = \frac{Mx}{ax-M}$, se verifica

$$\phi(x, y) \leq \frac{M}{a} \left(\frac{a}{2} - xy \right) \leq \frac{M}{a} \frac{a}{2} = \frac{\phi(\sqrt{a/6}, \sqrt{a/6})}{2}$$

es decir, en ellos, ϕ vale menos que en $(x^*, y^*) = (\sqrt{a/6}, \sqrt{a/6})$, punto que verifica

$$x^*y^* < a/2, \quad \frac{M}{a} < \frac{x^*y^*}{x^* + y^*}$$

En consecuencia, para determinar el valor máximo de ϕ en

$$\{(x, y) \in \mathbb{R}_{++}^2; xy \leq a/2\}$$

basta considerar ϕ en el conjunto cerrado y acotado del primer cuadrante delimitado por las curvas

$$xy = a/2 \text{ e } y = \frac{Mx}{ax - M}$$

Como en la parte de la frontera sobre $xy = a/2$ vale cero y en la parte que está sobre $y = \frac{Mx}{ax - M}$ vale menos que en $(\sqrt{a/6}, \sqrt{a/6})$ —punto que pertenece al interior—, ϕ ha de alcanzar su valor máximo en un punto crítico del interior de ese conjunto. Pero el único punto crítico de ϕ es precisamente $(x^*, y^*) = (\sqrt{a/6}, \sqrt{a/6})$, como sabemos. Así, pues, el cubo de arista $\sqrt{a/6}$ es realmente el paralelepípedo de mayor volumen que se puede construir con una superficie dada a .

En el caso general de una función de n variables y m restricciones, $m < n$,

$$\text{opt } f(x_1, \dots, x_n), \quad (x_1, \dots, x_n) \in S$$

s.a.

$$\begin{cases} g_1(x_1, \dots, x_n) = 0 \\ g_2(x_1, \dots, x_n) = 0 \\ \vdots \\ g_m(x_1, \dots, x_n) = 0 \end{cases}$$

el hessiano orlado es el determinante de orden $n + m$

$$|\bar{H}| = \begin{vmatrix} 0 & 0 & \dots & 0 & \frac{\partial g_1}{\partial x_1} & \frac{\partial g_1}{\partial x_2} & \dots & \frac{\partial g_1}{\partial x_n} \\ 0 & 0 & \dots & 0 & \frac{\partial g_2}{\partial x_1} & \frac{\partial g_2}{\partial x_2} & \dots & \frac{\partial g_2}{\partial x_n} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 & \frac{\partial g_m}{\partial x_1} & \frac{\partial g_m}{\partial x_2} & \dots & \frac{\partial g_m}{\partial x_n} \\ \frac{\partial g_1}{\partial x_1} & \frac{\partial g_1}{\partial x_2} & \dots & \frac{\partial g_1}{\partial x_n} & L_{x_1 x_1} & L_{x_1 x_2} & \dots & L_{x_1 x_n} \\ \frac{\partial g_1}{\partial x_2} & \frac{\partial g_1}{\partial x_2} & \dots & \frac{\partial g_1}{\partial x_n} & L_{x_2 x_1} & L_{x_2 x_2} & \dots & L_{x_2 x_n} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \frac{\partial g_m}{\partial x_1} & \frac{\partial g_m}{\partial x_2} & \dots & \frac{\partial g_m}{\partial x_n} & L_{x_n x_1} & L_{x_n x_2} & \dots & L_{x_n x_n} \end{vmatrix}$$

Denotando por $|\bar{H}_i|$ al menor

$$\begin{vmatrix} 0 & \dots & 0 & \frac{\partial g_1}{\partial x_1} & \dots & \frac{\partial g_1}{\partial x_i} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & 0 & \frac{\partial g_m}{\partial x_1} & \dots & \frac{\partial g_m}{\partial x_i} \\ \frac{\partial g_1}{\partial x_1} & \dots & \frac{\partial g_1}{\partial x_i} & L_{x_1 x_1} & \dots & L_{x_1 x_i} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \frac{\partial g_m}{\partial x_1} & \dots & \frac{\partial g_m}{\partial x_i} & L_{x_i x_1} & \dots & L_{x_i x_i} \end{vmatrix}$$

(o sea, el menor principal de $|\bar{H}|$ que contiene al menor principal de orden i del hessiano de L respecto a x_1, \dots, x_n) se puede demostrar en términos de los $n - m$ menores

$$|\bar{H}_{m+1}|, |\bar{H}_{m+2}|, \dots, |\bar{H}_n| \quad (= |\bar{H}|)$$

la siguiente condición suficiente de segundo orden para los puntos estacionarios x^* que hayamos obtenido al resolver el sistema de puntos críticos del lagrangiano:

- (i) Si todos los menores $|\bar{H}_{m+1}|, |\bar{H}_{m+2}|, \dots, |\bar{H}_n|$ tienen el mismo signo que $(-1)^m$, x^* es un mínimo local (estricto) de f restringida a $B = \{(x_1, \dots, x_n) \in S; g_j(x_1, \dots, x_n) = 0, j = 1, \dots, m\}$.
- (ii) Si el signo de $|\bar{H}_{m+1}|$ es el de $(-1)^{m+1}$ y los sucesivos menores $|\bar{H}_{m+2}|, \dots, |\bar{H}_n|$ van alternando de signo, x^* es un máximo local (estricto) de la función f restringida a $B = \{(x_1, \dots, x_n) \in S; g_j(x_1, \dots, x_n) = 0, j = 1, \dots, m\}$.

(EJERCICIO Formular estas condiciones para los casos $\{n = 3, m = 2\}$ y $\{n = 4, m = 2\}$.)

Véanse [1] y [2] para un tratamiento de esta cuestión. Hay que decir que las condiciones suficientes de segundo orden tienen, en general, una importancia limitada. Si se sabe que existe solución del problema de optimización, basta, para determinarla, comparar los valores de la función objetivo f en los puntos que satisfacen las condiciones de primer orden; las condiciones de segundo orden pueden simplemente servir para aligerar algo el trabajo descartando, por ejemplo, en un problema de maximización, los puntos que hayamos identificado como mínimos locales. Si no se conoce a priori la existencia de solución, las condiciones de segundo orden sólo permiten detectar óptimos locales, y esto, en muchas aplicaciones, tiene un interés limitado. Es cierto que, en determinadas circunstancias, como en el ejemplo anterior del paralelepípedo de volumen máximo, la "naturaleza del problema" puede aportar argumentos suplementarios que hagan plausible que un óptimo local lo sea también global, pero la demostración rigurosa puede ser, como también se aprecia en este ejemplo tan sencillo, bastante complicada.

Significado de los multiplicadores de Lagrange

En la resolución de problemas de optimización con restricciones de igualdad mediante el método de los multiplicadores de Lagrange, éstos desempeñan un papel meramente auxiliar en la obtención de los óptimos del problema:

intervienen en los cálculos y después nos “olvidamos” de ellos. Sin embargo, tienen, como vamos a ver, un significado que en ocasiones, en particular en Economía, los hace incluso más interesantes que los propios valores x^* del óptimo. Como en el ejemplo con que iniciamos este apartado, la restricción depende frecuentemente de un parámetro que describe un nivel o tope de la misma que puede variar, planteándose entonces el problema de optimización —en su versión más simple— en la forma

$$\begin{cases} \text{opt } f(x, y), (x, y) \in S \\ \text{s.a. } g(x, y) = c \end{cases}$$

Supongamos que este problema tiene solución para cada c de un cierto conjunto de \mathbb{R} . Las coordenadas del óptimo y el correspondiente valor del multiplicador de Lagrange dependerán, lógicamente, del parámetro c : $(x^*(c), y^*(c), \lambda^*(c))$, y, a través de aquéllas, lo mismo ocurrirá para el valor óptimo de la función objetivo: $f^*(c) = f((x^*(c), y^*(c)))$. Pues bien, bajo condiciones muy generales, se tiene el siguiente resultado:

$$\frac{df^*}{dc} = \lambda^*(c) \quad (6.61)$$

Su significado es que el multiplicador de Lagrange mide la *sensibilidad* del valor óptimo de f frente a variaciones en el nivel c de la restricción. Si recordamos la aproximación

$$\frac{df^*}{dc} \cong \frac{\Delta f^*}{\Delta c}$$

vemos que, en un determinado nivel c , $\lambda^*(c)$ representa —aproximadamente— el incremento que experimenta el valor óptimo de f cuando el nivel de la restricción pasa de valer c a valer $c + 1$. Como ilustración, resolvamos por el método de Lagrange el ejemplo citado:

$$\begin{aligned} \max \quad & \{x^2y\}, (x, y) \in \{(x, y) \in \mathbb{R}^2; x > 0, y > 0\} \\ \text{s.a.} \quad & x^2 + 4xy - a = 0 \end{aligned}$$

Lagrangiano:

$$L(x, y, \lambda) = x^2y - \lambda(x^2 + 4xy - a)$$

Puntos críticos del lagrangiano:

$$\begin{cases} \frac{\partial L}{\partial x} = 2xy - \lambda(2x + 4y) = 0 \\ \frac{\partial L}{\partial y} = x^2 - 4\lambda x = 0 \\ \frac{\partial L}{\partial \lambda} = -(x^2 + 4xy - a) = 0 \end{cases}$$

Teniendo en cuenta que ha de ser $x > 0, y > 0$, resulta la única solución $(x^*(a), y^*(a)) = (\sqrt{a/3}, (1/2)\sqrt{a/3})$ (que ya conocíamos) con multiplicador $\lambda^*(a) = (1/4)\sqrt{a/3}$. El valor en ella de f es

$$f^*(a) = f(x^*(a), y^*(a)) = \frac{a}{3} \cdot \frac{1}{2} \sqrt{\frac{a}{3}} = \frac{a}{6} \sqrt{\frac{a}{3}}$$

que, por los argumentos ya vistos, es efectivamente el valor máximo de f requerido. Vemos que efectivamente

$$\frac{df^*}{da} = \frac{1}{4} \sqrt{\frac{a}{3}} = \lambda^*(a)$$

Si, por ejemplo, se dispone de 12 m^2 de plancha de acero, el depósito que se puede construir tendrá un volumen de $f^*(12) = 4 \text{ m}^3$. Si se dispusiese de 1 m^2 más de chapa, el volumen del depósito aumentaría aproximadamente en $\lambda^*(12) = 0,5 \text{ m}^3$.

Veamos el resultado (6.61). Se tiene, con precisión, el siguiente

Teorema 6.14 Dadas dos funciones f y g de clase C^2 en un conjunto abierto $S \subseteq \mathbb{R}^2$, se considera el problema

$$\begin{aligned} \text{opt } f(x, y), \quad (x, y) \in S \\ \text{s.a. } g(x, y) = c \end{aligned} \quad (6.62)$$

Supongamos que para $c = c_0$ el problema (6.62) tiene una solución local (x_0^*, y_0^*) en la que se verifican las condiciones necesarias de primer orden así como, junto con su multiplicador asociado λ_0^* , las condiciones suficientes de segundo orden de óptimo local estricto. Entonces, en un cierto entorno I de c_0 , están definidas unas funciones $x^*(c), y^*(c), \lambda^*(c)$ de clase C^1 tales que $x^*(c_0) = x_0^*, y^*(c_0) = y_0^*, \lambda^*(c_0) = \lambda_0^*$ y $(x^*(c), y^*(c))$ es óptimo local de (6.62) (el único en un cierto entorno de (x_0^*, y_0^*)) con multiplicador $\lambda^*(c)$. Además, poniendo $f^*(c) = f(x^*(c), y^*(c))$, se tiene

$$\frac{df^*(c)}{dc} = \lambda^*(c) \quad (6.63)$$

Demostración. El lagrangiano del problema (6.62) es

$$L(x, y, \lambda) = f(x, y) - \lambda[g(x, y) - c]$$

Las condiciones necesarias de primer orden son

$$\begin{aligned}\frac{\partial L}{\partial x}(x, y, \lambda) &= \frac{\partial f}{\partial x}(x, y) - \lambda \frac{\partial g}{\partial x}(x, y) = 0 \\ \frac{\partial L}{\partial y}(x, y, \lambda) &= \frac{\partial f}{\partial y}(x, y) - \lambda \frac{\partial g}{\partial y}(x, y) = 0 \\ \frac{\partial L}{\partial \lambda}(x, y, \lambda) &= -[g(x, y) - c] = 0\end{aligned}\tag{6.64}$$

Considerando (6.64) como un sistema en las variables (x, y, λ, c) , vemos que, por hipótesis, $(x_0^*, y_0^*, \lambda_0^*, c_0)$ lo satisface y que el jacobiano en él de las funciones que definen el sistema respecto a (x, y, λ) :

$$\begin{vmatrix} L_{xx} & L_{xy} & -g_x \\ L_{yx} & L_{yy} & -g_y \\ -g_x & -g_y & 0 \end{vmatrix} = \begin{vmatrix} 0 & g_x & g_y \\ g_x & L_{xx} & L_{xy} \\ g_y & L_{yx} & L_{yy} \end{vmatrix} = |\hat{H}(x_0^*, y_0^*, \lambda_0^*)|\tag{6.65}$$

es distinto de cero (verificación de la condición suficiente de segundo orden). En consecuencia, por el teorema de la función implícita, existen $\delta > 0$, $\varepsilon > 0$ tales que para cada $c \in (c_0 - \delta, c_0 + \delta)$, el sistema (6.64) tiene única solución $x^*(c)$, $y^*(c)$, $\lambda^*(c)$ que verifica $|x^*(c) - x_0^*| < \varepsilon$, $|y^*(c) - y_0^*| < \varepsilon$, $|\lambda^*(c) - \lambda_0^*| < \varepsilon$ y las funciones $x^*(c)$, $y^*(c)$, $\lambda^*(c)$ son de clase C^1 en $(c_0 - \delta, c_0 + \delta)$. Por otra parte, el hessiano orlado (6.65) es distinto de cero (y mantiene su signo) en $(x^*(c), y^*(c), \lambda^*(c))$ para todo c de un cierto entorno $I \subseteq (c_0 - \delta, c_0 + \delta)$ de c_0 , por la continuidad de las funciones que lo constituyen. Se tiene, por tanto, que $(x^*(c), y^*(c))$, junto con $\lambda^*(c)$, satisface las condiciones necesarias de primer orden (6.64) y las condiciones suficientes de segundo orden, por lo que, por el teorema 6.13, $(x^*(c), y^*(c))$ es un óptimo local del problema (6.62) con multiplicador $\lambda^*(c)$.

Para demostrar (6.63), consideremos la función compuesta

$$c \mapsto L(x^*(c), y^*(c), \lambda^*(c)) = f(x^*(c), y^*(c)) - \lambda^*(c)[g(x^*(c), y^*(c)) - c]$$

Derivando respecto a c por la regla de la cadena:

$$\begin{aligned} \frac{dL}{dc} &= \frac{\partial f}{\partial x} \frac{dx^*(c)}{dc} + \frac{\partial f}{\partial y} \frac{dy^*(c)}{dc} - \frac{d\lambda^*(c)}{dc} [g(x^*(c), y^*(c)) - c] - \\ &\quad - \lambda^*(c) \left[\frac{\partial g}{\partial x} \frac{dx^*(c)}{dc} + \frac{\partial g}{\partial y} \frac{dy^*(c)}{dc} - 1 \right] = \\ &= \frac{dx^*(c)}{dc} \left[\frac{\partial f}{\partial x} - \lambda^*(c) \frac{\partial g}{\partial x} \right] + \frac{dy^*(c)}{dc} \left[\frac{\partial f}{\partial y} - \lambda^*(c) \frac{\partial g}{\partial y} \right] + \\ &\quad - \frac{d\lambda^*(c)}{dc} [g(x^*(c), y^*(c)) - c] + \lambda^*(c) = \lambda^*(c) \end{aligned}$$

ya que los tres primeros sumandos se anulan por la verificación de las condiciones necesarias (6.64) en cada óptimo $(x^*(c), y^*(c), \lambda^*(c))$. Como, por otra parte, $L(x^*(c), y^*(c), \lambda^*(c)) \equiv f(x^*(c), y^*(c))$, pues L y f coinciden sobre el conjunto factible del problema de optimización, se tendrá

$$\frac{df(x^*(c), y^*(c))}{dc} = \lambda^*(c)$$

En el caso general de una función de n variables sujeta a m restricciones, $m < n$, el resultado anterior toma la forma siguiente:

Teorema 6.15 Dadas las funciones f, g_1, \dots, g_m de clase C^2 en un conjunto abierto $S \subseteq \mathbb{R}^n$, se considera el problema

$$\begin{aligned} \text{opt } f(x_1, \dots, x_n), \quad (x_1, \dots, x_n) \in S \\ \text{s. a.} \\ \begin{cases} g_1(x_1, \dots, x_n) = c_1 \\ g_2(x_1, \dots, x_n) = c_2 \\ \vdots \\ g_m(x_1, \dots, x_n) = c_m \end{cases} \end{aligned} \quad (6.66)$$

Supongamos que para $c = c_0 = (c_{10}, \dots, c_{m0}) \in \mathbb{R}^m$ el problema (6.66) tiene una solución local $x_0^* = (x_{10}^*, \dots, x_{n0}^*)$ en la que se verifican las condiciones necesarias de primer orden así como, junto con su multiplicador asociado $\lambda_0^* = (\lambda_{10}^*, \dots, \lambda_{m0}^*)$ las condiciones suficientes de segundo orden de óptimo local estricto. Entonces, para $c = (c_1, \dots, c_m)$ moviéndose en un cierto entorno U de c_0 , están definidas unas funciones $x^*(c) = (x_1^*(c), \dots, x_n^*(c))$, $\lambda^*(c) = (\lambda_1^*(c), \dots, \lambda_m^*(c))$ de clase C^1 tales que $x^*(c_0) = x_0^*$, $\lambda^*(c_0) = \lambda_0^*$ y

$x^*(c)$ es óptimo local de (6.66) (el único en un cierto entorno de x_0^*) con multiplicadores $\lambda^*(c)$. Además, poniendo $f^*(c) = f(x^*(c))$, se tiene

$$\nabla f^*(c) = \lambda^*(c) \quad (6.67)$$

es decir

$$\frac{\partial f^*}{\partial c_j}(c_1, \dots, c_m) = \lambda_j^*(c_1, \dots, c_m)$$

Si se supone que efectivamente el problema (6.66) tiene solución para todo c de un cierto entorno de c_0 y que la función f^* es de clase C^1 , entonces la demostración por la regla de la cadena de (6.67) es bastante simple (inténtese como ejercicio).

NOTAS. 1. Como consecuencia directa del resultado anterior se deduce que la función de valor óptimo $f^*(c)$ es en realidad de clase C^2 , pues $\nabla f^*(c) = \lambda^*(c)$ para todo $c \in U$ y $\lambda^*(c)$ es de clase C^1 .

2. El teorema 6.15 admite una generalización aún mayor si se supone que el vector de parámetros $c = (c_1, \dots, c_m)$ afecta no sólo a los términos independientes de las restricciones, sino a las propias funciones g_1, \dots, g_m que las definen e incluso a la función objetivo f . En este caso, considerando el problema

$$\begin{array}{ll} \text{opt } f(x, c), & x = (x_1, \dots, x_n) \in S \\ \text{s.a.} & \\ & \begin{cases} g_1(x, c) = 0 \\ g_2(x, c) = 0 \\ \vdots \\ g_m(x, c) = 0 \end{cases} \end{array}$$

y análogas hipótesis que en el teorema 6.15, se concluye que las funciones $x^*(c)$ y $\lambda^*(c)$ están definidas en un cierto entorno U de c_0 y son también de clase C^1 . Además, la función de valor óptimo es de clase C^2 en U y verifica

$$\frac{\partial f^*}{\partial c_j}(c) = \frac{\partial L}{\partial c_j}(x^*(c), \lambda^*(c), c), \quad j = 1, \dots, m$$

siendo $L(x, \lambda, c)$ el lagrangiano asociado al problema de partida. Este resultado se conoce con el nombre de *teorema de la envolvente*.

Clasificación de una forma cuadrática mediante sus autovalores

Sea $q(x) = x^T A x = \langle x, A x \rangle$ con $A = (a_{ij})$ una matriz simétrica. Nos planteamos el problema

$$\begin{array}{ll} \text{opt } q(x) & \\ \text{s.a. } x_1^2 + \dots + x_n^2 = 1 & \end{array} \quad (6.68)$$

La esfera unidad $S_1 = \{x \in \mathbb{R}^n; x_1^2 + \dots + x_n^2 = 1\}$ es un conjunto cerrado y acotado de \mathbb{R}^n , por lo que $q(x)$ alcanza en él su valor máximo y su valor mínimo. Para determinarlos utilizamos el método de Lagrange:

Lagrangiano:

$$L(x, \lambda) = \langle x, Ax \rangle - \lambda(\langle x, x \rangle - 1) = \langle x, (A - \lambda I)x \rangle + \lambda$$

Puntos estacionarios:

$$\begin{cases} \frac{\partial L}{\partial x_i} = \langle e_i, (A - \lambda I)x \rangle + \langle x, (A - \lambda I)e_i \rangle = 0 \\ |x| = 1 \end{cases} \quad (6.69)$$

Utilizando la simetría del producto escalar y el hecho de que ser A simétrica significa que $\langle x, Ay \rangle = \langle Ax, y \rangle$ para todo par $x, y \in \mathbb{R}^n$, se deduce que las n primeras condiciones de (6.69) son

$$\langle (A - \lambda I)x, e_i \rangle = 0, \quad i = 1, \dots, n$$

lo que implica

$$(A - \lambda I)x = 0$$

Así, pues, en los óptimos $x \in S_1$ ha de verificarse

$$Ax = \lambda x$$

es decir, que x es *autovector* de A con *autovalor* asociado λ . Además

$$q(x) = \langle x, Ax \rangle = \langle x, \lambda x \rangle = \lambda \langle x, x \rangle = \lambda$$

y, por tanto, el valor mínimo de $q(x)$ en S_1 es el menor autovalor λ_m de A y su valor máximo, el mayor autovalor λ_M de A . Se tiene, pues,

$$\lambda_m \leq q(x) \leq \lambda_M \quad \text{para todo } x \in S_1$$

Si $x \in \mathbb{R}^n$, $x \neq 0$, entonces $x/|x| \in S_1$, de donde

$$\lambda_m \leq q\left(\frac{x}{|x|}\right) \leq \lambda_M$$

y, por ser q homogénea de grado 2:

$$q\left(\frac{x}{|x|}\right) = \frac{1}{|x|^2} q(x), \quad (6.70)$$

$$\lambda_m |x|^2 \leq q(x) \leq \lambda_M |x|^2 \quad \text{para todo } x \in \mathbb{R}^n \quad (6.71)$$

Dedúzcase de (6.71), junto con (6.70) y la propiedad $q(x) = \lambda |x|^2$ si x es autovector de A , la clasificación de $q(x)$ en términos de los autovalores de A .

6.3 Restricciones de desigualdad

En el apartado anterior hemos estudiado lo que supone para la toma de decisión del "optimizador" el que las variables de decisión x_1, \dots, x_n estén sujetas a restricciones de la forma $g(x_1, \dots, x_n) = c$. En ocasiones, es más natural que dichas restricciones se planteen en forma de *desigualdades*: $g(x_1, \dots, x_n) \leq c$; por ejemplo, en un proceso productivo, $g(x)$ puede representar el coste total del mismo y, si se dispone de un presupuesto de c unidades, puede que no sea forzoso agotarlo para lograr el objetivo óptimo propuesto. El problema se plantearía entonces como el de optimizar la función objetivo $f(x_1, \dots, x_n)$ (la que corresponda al objetivo propuesto) sujeta a la restricción $g(x_1, \dots, x_n) \leq c$. Otras veces, las *restricciones de desigualdad* aparecen como restricciones sobre los valores que pueden tomar las variables de decisión x_1, \dots, x_n .

El ejemplo 3 del apartado 1.3:

$$\begin{cases} \max \{4x_1 + 3x_2\} \\ \text{s.a.} \\ x_1 \leq 10 \\ x_2 \leq 12 \\ 2x_1 + x_2 \leq 24 \\ x_1, x_2 \geq 0 \end{cases} \quad (6.72)$$

es una buena muestra de este tipo de *problemas de optimización con restricciones de desigualdad*. Se trata, como dijimos allí, de un *programa lineal* porque tanto la función objetivo como las funciones que definen las restricciones son funciones lineales.

En el apartado 3.3 nos planteábamos, para ilustrar la aplicación de la diferenciabilidad al cálculo aproximado, un proceso productivo en el que la *función de producción* era la función de Cobb-Douglas

$$f(x, y) = 30x^{1/3}y^{2/3} \quad (6.73)$$

donde x e y representan, respectivamente, las unidades de capital (maquinaria) y de trabajo. Un problema de gran interés económico para una empresa es el de *minimización de costes*; su resolución le proporcionará información valiosa sobre la estructura de sus costes de producción. En dicho problema, la empresa (para la que se supone que se conoce la función de producción f , reflejo de su "capacidad tecnológica" para transformar en producto diversos tipos de *inputs*) ha de hacer frente a unos costes unitarios p y q de los *factores* x e y (si se consideran únicamente dos clases de *inputs*), y a un nivel mínimo de *producción* prescrito c . El problema es, por tanto,

$$\begin{cases} \min \{px + qy\} \\ \text{s.a. } f(x, y) \geq c \end{cases} \quad (6.74)$$

Adoptaremos para el estudio de este tipo de problemas la siguiente formulación general:

$$\begin{array}{ll} \text{opt } f(x_1, \dots, x_n), & (x_1, \dots, x_n) \in S \subseteq \text{Dom } f \\ \text{s.a.} & \\ \left\{ \begin{array}{l} g_1(x_1, \dots, x_n) \leq 0 \\ g_2(x_1, \dots, x_n) \leq 0 \\ \vdots \\ g_m(x_1, \dots, x_n) \leq 0 \end{array} \right. & \end{array} \quad (6.75)$$

El conjunto de **soluciones factibles** (o **conjunto factible**) es

$$B = \{(x_1, \dots, x_n) \in S; g_j(x_1, \dots, x_n) = 0, j = 1, \dots, m\} \quad (6.76)$$

Observaciones:

1. Bastaría estudiar únicamente bien el problema de minimización bien el de maximización, ya que maximizar $f(x)$ equivale a minimizar $\{-f(x)\}$ y, así, todo problema de maximización se puede convertir en uno de minimización y viceversa.
2. Las restricciones se suponen de la forma

$$g_j(x) \leq 0 \quad (6.77)$$

Si una restricción apareciese en la forma $g_j(x) \leq c$, bastaría escribir

$$g_j(x) - c \leq 0$$

Por otro lado, cualquier restricción de la forma $g_j(x) \geq 0$ se puede convertir en una de la forma (6.77) poniendo

$$-g_j(x) \leq 0$$

3. El número m de restricciones puede ser cualquiera; aquí no se aplican las consideraciones que hacíamos para las restricciones de igualdad y no hay la limitación $m < n$.

Para problemas con dos variables de decisión y funciones f, g_j no muy complicadas, pueden ser aplicables métodos gráficos de resolución, como comentábamos en el apartado 1.3. Pero estamos interesados, lógicamente, en establecer métodos analíticos de resolución que sean aplicables al problema general (6.75), del cual hay que decir que aún se está lejos de una resolución completamente satisfactoria. Nuestro objetivo es establecer, como en los dos

apartados anteriores, *condiciones necesarias* para que el problema (6.75) alcance un óptimo (local) en un punto $x = x^*$. Comencemos con la situación más simple consistente en optimizar —concretamente, minimizar, para fijar las ideas— una función objetivo de dos variables sujeta a una sola restricción:

$$\begin{cases} \min f(x_1, x_2), & (x_1, x_2) \in S \\ \text{s.a. } g(x_1, x_2) \leq 0 \end{cases} \quad (6.78)$$

Suponemos que f y g son dos funciones de clase C^1 en el conjunto abierto $S \subseteq \mathbb{R}^2$. Como hemos visto en el capítulo 4, los puntos (x_1, x_2) que satisfacen $g(x_1, x_2) = 0$ forman, en general, una curva en el plano que representa la frontera entre la región en la que $g < 0$ y la región en que $g > 0$ (figura 6.20).

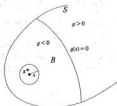


Figura 6.20

Supongamos que el óptimo se produce en un punto $x^* = (x_1^*, x_2^*)$ en el que $g < 0$ (figura 6.20). Entonces, por la continuidad de g , existe $\varepsilon < 0$ tal que $g(x) < 0$ para todo $x \in B_\varepsilon(x^*)$ (o sea, x^* es un punto interior de B), $B_\varepsilon(x^*) \subset S$ por ser S abierto, y, por hipótesis, $f(x) \geq f(x^*)$ para todo $x \in B_\varepsilon(x^*)$. En consecuencia, x^* es un punto crítico de la función de la función $f(x_1, x_2)$ y en él ha de verificarse

$$\frac{\partial f}{\partial x_1} = \frac{\partial f}{\partial x_2} = 0 \quad (6.79)$$

(y lo mismo ocurriría para un máximo). La *condición necesaria de primer orden* sería, por tanto, la de los problemas de óptimos sin restricciones: la restricción definida por $g(x_1, x_2)$ no influiría, en realidad, en la determinación del óptimo del problema (en un entorno de él también se tendría $g < 0$, luego esa restricción no es “propia” del óptimo x^* ; en un problema de Economía como el aludido al principio del apartado, cabría decir que si un recurso limitado, como el del presupuesto, no se agota para lograr el óptimo, la limitación que supone no tiene ninguna intervención activa en la determinación de ese óptimo).

Si el mínimo se produce en un punto x^* en el que $g(x^*) = 0$ (en cuyo caso se dice que la restricción está *saturada*), entonces x^* será solución del problema

$$\begin{cases} \min f(x_1, x_2), & (x_1, x_2) \in S \\ \text{s.a. } g(x_1, x_2) = 0 \end{cases} \quad (6.80)$$

Si suponemos que $\nabla g(x_1^*, x_2^*) \neq (0, 0)$ (lo que, en particular, implica que $g(x_1, x_2) = 0$ define localmente una curva en cada uno de cuyos lados g toma signos opuestos y, en consecuencia, que (x_1^*, x_2^*) es un *punto frontera* del conjunto factible B que pertenece a S), entonces, como sabemos del apartado anterior, existe un número λ^* tal que

$$\nabla f(x_1^*, x_2^*) = \lambda^* \nabla g(x_1^*, x_2^*) \quad (6.81)$$

Esta es, en este caso, la *condición necesaria de primer orden* de óptimo local. Afirmamos que el *multiplicador* λ^* ha de ser necesariamente ≤ 0 (y ≥ 0 para el correspondiente problema de maximización). Para convencernos de ello, argumentemos geoméricamente de un modo intuitivo (figura 6.21; a continuación daremos la demostración analítica rigurosa).

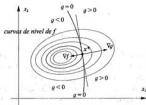


Figura 6.21

(6.81) significa que ∇f y ∇g son colineales en x^* ; si fuese $\lambda^* > 0$, entonces $-\nabla f (= \lambda^*(-\nabla g))$ apuntaría hacia el conjunto factible, pues $-\nabla g$ apunta hacia $g < 0$ al señalar el sentido de decrecimiento —más rápido— de g . Como, a su vez, $-\nabla f$ indica la dirección y sentido de decrecimiento más rápido sobre la superficie $z = f(x_1, x_2)$, o sea, de la función f , eso implicaría que hay puntos en el interior del conjunto factible en los que f valdría *menos* que lo que vale en x^* , es decir, éste no sería solución del problema de minimización contra la hipótesis. Así, pues, ha de ser $\lambda^* \leq 0$ (si $\lambda^* = 0$, entonces $\nabla f(x_1^*, x_2^*) = (0, 0)$ y se trataría de un punto crítico de f que está en la frontera del conjunto factible). Argumentando de manera análoga, veríamos que ha de ser $\lambda^* \geq 0$ si el problema es de maximización.

Justifiquemos este resultado analíticamente. Podemos suponer, sin pérdida de generalidad, que la hipótesis $\nabla g(x_1^*, x_2^*) \neq (0, 0)$ se traduce en que

$$\frac{\partial g}{\partial x_2}(x_1^*, x_2^*) \neq 0 \quad (6.82)$$

Consideremos la ecuación

$$g(x_1^*, x_2) + \alpha = 0 \quad (6.83)$$

en las variables α, x_2 ; $\{\alpha = 0, x_2 = x_2^*\}$ es solución de (6.83). Por el teorema de la función implícita, (6.82) implica que para α suficientemente pequeño está definida una única función $x_2 = \varphi(\alpha)$ tal que $x_2^* = \varphi(0)$ y $g(x_1^*, \varphi(\alpha)) + \alpha = 0$. Para los valores $\alpha > 0$, los puntos $(x_1^*, \varphi(\alpha))$ pertenecen al conjunto factible B pues $g(x_1^*, \varphi(\alpha)) = -\alpha < 0$, por lo que, por hipótesis,

$$f(x_1^*, x_2^*) \leq f(x_1^*, \varphi(\alpha))$$

y, en consecuencia,

$$\begin{aligned} 0 &\leq \frac{f(x_1^*, \varphi(\alpha)) - f(x_1^*, x_2^*)}{\alpha} = \frac{\frac{\partial f}{\partial x_2}(x_1^*, x_2^*) \cdot (\varphi(\alpha) - x_2^*) + o(\varphi(\alpha) - x_2^*)}{\alpha} = \\ &= (\varphi(0) = x_2^*) = \frac{\partial f}{\partial x_2}(x_1^*, x_2^*) \cdot \frac{\varphi(\alpha) - \varphi(0)}{\alpha} + \frac{o(\varphi(\alpha) - \varphi(0))}{\varphi(\alpha) - \varphi(0)} \cdot \frac{\varphi(\alpha) - \varphi(0)}{\alpha} \\ &\xrightarrow{\alpha \rightarrow 0^+} \frac{\partial f}{\partial x_2}(x_1^*, x_2^*) \cdot \varphi'(0) + 0 \cdot \varphi'(0) = \frac{\partial f}{\partial x_2}(x_1^*, x_2^*) \cdot \frac{-1}{(\partial g / \partial x_2)(x_1^*, x_2^*)} = \\ &= (\text{por (6.81)}) = -\lambda^* \end{aligned}$$

Así pues, $0 \leq -\lambda^*$, o sea, $\lambda^* \leq 0$.

Naturalmente, en un problema concreto no sabemos a priori si el óptimo se produce en la frontera o en el interior de B , es decir, si la restricción $g \leq 0$ está o no saturada en él. Conviene entonces reunir la alternativa anterior en una condición necesaria de primer orden que permita en la práctica la búsqueda de los óptimos:

Teorema 6.16 (Kuhn-Tucker) Sean f y g dos funciones de clase C^1 en un conjunto abierto $S \subseteq \mathbb{R}^2$. Supongamos que $x^* = (x_1^*, x_2^*)$ es una solución (local) del problema

$$\begin{aligned} \min (\text{resp. } \max) \quad & f(x_1, x_2), \quad (x_1, x_2) \in S \\ \text{s.a.} \quad & g(x_1, x_2) \leq 0 \end{aligned} \quad (6.84)$$

y supongamos también que, si $g(x_1^*, x_2^*) = 0$, $\nabla g(x_1^*, x_2^*) \neq (0, 0)$. Entonces existe un número λ^* tal que

$$\begin{aligned} \nabla f(x_1^*, x_2^*) &= \lambda^* \nabla g(x_1^*, x_2^*), \\ \lambda^* &\leq 0 \text{ (resp. } \geq 0) \text{ y } \lambda^* \cdot g(x_1^*, x_2^*) = 0. \end{aligned} \quad (6.85)$$

(6.85) se conoce como *sistema de condiciones de Kuhn-Tucker* del problema (6.84). Obsérvese que, efectivamente, en las condiciones (6.85) está la alternativa:

- Si $g(x_1^*, x_2^*) < 0$, entonces $\lambda^* = 0$ (pues ha de ser $\lambda \cdot g = 0$) y, por tanto,

$$\frac{\partial f}{\partial x_1}(x_1^*, x_2^*) = \frac{\partial f}{\partial x_2}(x_1^*, x_2^*) = 0$$

es decir, (x_1^*, x_2^*) es punto crítico de f .

- Si $g(x_1^*, x_2^*) = 0$, entonces se verifica $\nabla f(x_1^*, x_2^*) = \lambda^* \nabla g(x_1^*, x_2^*)$ con $\lambda^* \leq 0$ (resp. $\lambda^* \geq 0$).

En la práctica se procede como sigue:

Se forma (como en el caso de restricciones de igualdad) el *lagrangiano*

$$L(x_1, x_2, \lambda) = f(x_1, x_2) - \lambda g(x_1, x_2)$$

y se resuelve el *sistema de condiciones de Kuhn-Tucker*:

$$\begin{cases} \frac{\partial L}{\partial x_1} = \frac{\partial f}{\partial x_1} - \lambda \frac{\partial g}{\partial x_1} = 0 \\ \frac{\partial L}{\partial x_2} = \frac{\partial f}{\partial x_2} - \lambda \frac{\partial g}{\partial x_2} = 0 \\ g(x_1, x_2) \leq 0, \lambda \leq 0 \text{ (resp. } \lambda \geq 0), \lambda \cdot g(x_1, x_2) = 0 \end{cases} \quad (6.86)$$

Entre sus soluciones $(x_1^*, x_2^*, \lambda^*)$ (quedándonos con las dos primeras coordenadas (x_1^*, x_2^*)) ha(n) de estar, si existe(n), la(s) solución(es) del problema (6.84).

EjemPlo 17. Se trata de resolver el problema

$$\begin{cases} \min \{x^3 - y^2\} \\ \text{s.a. } x^2 + y^2 \leq 1 \end{cases} \quad (6.87)$$

El conjunto factible B es la bola unidad cerrada de \mathbb{R}^2 , con lo que el problema tiene solución. Para determinarla(s), utilizamos el método expuesto.

Lagrangiano: $L(x, y, \lambda) = x^3 - y^2 - \lambda(x^2 + y^2 - 1)$

Condiciones de Kuhn-Tucker:

$$\begin{cases} \frac{\partial L}{\partial x} = 3x^2 - 2\lambda x = 0 \\ \frac{\partial L}{\partial y} = -2y - 2\lambda y = 0 \\ x^2 + y^2 - 1 \leq 0, \lambda \leq 0, \lambda(x^2 + y^2 - 1) = 0 \end{cases} \quad (6.88)$$

Alternativa:

i) g no se satura (o sea, estamos en el interior de B): $g(x, y) < 0$ y $\lambda = 0$. (6.88) es

$$\begin{cases} \frac{\partial f}{\partial x} = 3x^2 = 0 \\ \frac{\partial f}{\partial y} = -2y = 0 \end{cases}$$

sistema de puntos críticos de f . La única solución es $(0, 0)$, que sí satisface $g(x, y) < 0$. Por tanto, es un candidato a solución del problema (6.87).

ii) g sí se satura, o sea, nos situamos en la frontera $\{(x, y); x^2 + y^2 - 1 = 0\}$ de B . (6.88) toma la forma

$$\begin{cases} 3x^2 - 2\lambda x = 0 \\ -2y - 2\lambda y = 0 \\ x^2 + y^2 = 1, \lambda \leq 0 \end{cases}$$

o equivalentemente

$$\begin{cases} x(3x - 2\lambda) = 0 \\ y(1 + \lambda) = 0 \\ x^2 + y^2 = 1, \lambda \leq 0 \end{cases}$$

Comenzando con la alternativa $y = 0$ o bien $\lambda = -1$ que proporciona la segunda ecuación, obtenemos las soluciones de la tabla siguiente (independientemente del signo de λ):

x^*	y^*	λ^*	$f(x^*, y^*)$	
0	0	0	0	
1	0	3/2	1	máx
-1	0	-3/2	-1	mín
0	1	-1	-1	mín
0	-1	-1	-1	mín
-2/3	$\sqrt{5}/3$	-1	-23/27	
-2/3	$-\sqrt{5}/3$	-1	-23/27	

Cinco de estos puntos son candidatos a mínimo, uno de ellos a máximo y el punto $(0, 0)$, para el que $\lambda = 0$, cumple las condiciones necesarias tanto de mínimo como de máximo; nótese además que $\nabla g(x, y) = (2x, 2y)$ sólo se anula en $(0, 0)$, punto que no satura la restricción, con lo que en los seis puntos restantes se cumple la condición de regularidad $\nabla g(x, y) \neq (0, 0)$. Evaluando la función objetivo en todos los candidatos vemos que las soluciones del problema de minimización (6.87) son $(-1, 0)$, $(0, 1)$, $(0, -1)$, mientras que $(1, 0)$ lo sería para el correspondiente problema de maximización.

Generalizando fácilmente los argumentos anteriores, demostraremos en el capítulo 9 el siguiente resultado relativo al problema (6.75):

Teorema 6.17 (Kuhn-Tucker) Sean f, g_1, \dots, g_m , funciones de clase C^1 en un subconjunto abierto $S \subseteq \mathbb{R}^n$ con valores en \mathbb{R} . Supongamos que $x^* = (x_1^*, \dots, x_n^*)$ es un óptimo local de f en el conjunto

$$B = \{x \in S; g_j(x) \leq 0, j = 1, \dots, m\} \quad (6.89)$$

o sea, una solución local del problema (6.75). Reordenando las funciones g_j si es necesario, podemos suponer que las restricciones de desigualdad que se saturan en x^* son $g_1(x^*) = 0, \dots, g_r(x^*) = 0$, con $r \leq m$. Pues bien, si los vectores $\{\nabla g_1(x^*), \dots, \nabla g_r(x^*)\}$ son linealmente independientes, entonces existen constantes $\lambda_1^*, \dots, \lambda_m^*$ tales que

$$\begin{cases} \nabla f(x^*) = \lambda_1^* \nabla g_1(x^*) + \dots + \lambda_m^* \nabla g_m(x^*) \\ y, \text{ para todo } j = 1, \dots, m, \quad \lambda_j^* g_j(x^*) = 0 \text{ y} \\ \lambda_j^* \leq 0 \text{ si } x^* \text{ es un mínimo o } \lambda_j^* \geq 0 \text{ si } x^* \text{ es un máximo.} \end{cases} \quad (6.90)$$

En la práctica, el esquema de trabajo para intentar resolver el problema (6.75) es el mismo que en los problemas con restricciones de igualdad:

- (i) Se forma el lagrangiano del problema:

$$L(x, \lambda) = f(x) - \lambda_1 g_1(x) - \dots - \lambda_m g_m(x)$$

(ii) Se resuelve el sistema de condiciones de Kuhn-Tucker:

$$\begin{cases} \frac{\partial L}{\partial x_i} = \frac{\partial f}{\partial x_i} - \sum_{j=1}^m \lambda_j \frac{\partial g_j}{\partial x_i} = 0, \quad i = 1, \dots, n \\ g_j(x) \leq 0, \quad \lambda_j \leq 0 \text{ (o bien } \lambda_j \geq 0), \quad \lambda_j g_j(x) = 0, \quad j = 1, \dots, m \end{cases} \quad (6.91)$$

Sus soluciones $(x_1^*, \dots, x_n^*, \lambda_1^*, \dots, \lambda_m^*)$ (en las que los multiplicadores de Kuhn-Tucker asociados a las restricciones de desigualdad han de ser ≤ 0 para un problema de minimización y ≥ 0 para uno de maximización) proporcionan, quedándose con las n primeras coordenadas (x_1^*, \dots, x_n^*) , candidatos a soluciones del problema (6.75). Si éste tiene solución, ha de estar entre esos candidatos: se evalúa la función f en ellos y, si el problema es uno de minimización, el menor de los valores obtenidos da la solución y, si es de maximización, la da el mayor de ellos. Para dilucidar si efectivamente existe solución habrá que apelar a argumentos suplementarios, generalmente basados en el teorema de Weierstrass —como en el ejemplo anterior— o en propiedades de convexidad que veremos después. En la resolución de (6.91) conviene adoptar una estrategia combinatoria que tenga en cuenta la alternativa saturación/no saturación de cada una de las restricciones de desigualdad:

EJEMPLO 18. Resolvamos el problema

$$\begin{aligned} &\min\{x_1 + 2x_2\}, \quad (x_1, x_2) \in \mathbb{R}^2 \\ &\text{s.a.} \\ &\begin{cases} x_1 - 2x_2 \leq 2 \\ x_1^2 + x_2^2 \leq 4 \end{cases} \end{aligned} \quad (6.92)$$

El problema tiene solución ya que el conjunto factible

$$B = \{(x_1, x_2) \in \mathbb{R}^2; x_1 - 2x_2 \leq 2; x_1^2 + x_2^2 \leq 4\}$$

es un conjunto cerrado y acotado (inténtese una resolución gráfica del problema en paralelo a la resolución analítica). El lagrangiano es:

$$L(x_1, x_2, \lambda_1, \lambda_2) = x_1 + 2x_2 - \lambda_1(x_1 - 2x_2 - 2) - \lambda_2(x_1^2 + x_2^2 - 4)$$

Las condiciones de Kuhn-Tucker son:

$$\begin{cases} \frac{\partial L}{\partial x_1} = \frac{\partial f}{\partial x_1} - \lambda_1 \frac{\partial g_1}{\partial x_1} - \lambda_2 \frac{\partial g_2}{\partial x_1} = 1 - \lambda_1 - 2\lambda_2 x_1 = 0 \\ \frac{\partial L}{\partial x_2} = \frac{\partial f}{\partial x_2} - \lambda_1 \frac{\partial g_1}{\partial x_2} - \lambda_2 \frac{\partial g_2}{\partial x_2} = 2 + 2\lambda_1 - 2\lambda_2 x_2 = 0 \\ x_1 - 2x_2 - 2 \leq 0; \quad x_1^2 + x_2^2 - 4 \leq 0; \quad \lambda_1 \leq 0; \quad \lambda_2 \leq 0 \\ \lambda_1(x_1 - 2x_2 - 2) = 0; \quad \lambda_2(x_1^2 + x_2^2 - 4) = 0 \end{cases} \quad (6.93)$$

Caben 4 posibilidades:

$$\begin{array}{ll} \text{a) } g_1 < 0; \quad g_2 < 0 & \text{b) } g_1 = 0; \quad g_2 < 0 \\ \text{c) } g_1 < 0; \quad g_2 = 0 & \text{d) } g_1 = 0; \quad g_2 = 0 \end{array}$$

Veamos una por una:

a) $g_1(x_1, x_2) < 0$; $g_2(x_1, x_2) < 0$, o sea, estamos en el interior de B . Entonces, $\lambda_1 = 0$, $\lambda_2 = 0$ por la condición $\lambda_1 g_1 = 0$, $\lambda_2 g_2 = 0$. Las dos primeras ecuaciones de (6.93) son

$$\frac{\partial f}{\partial x_1} = 0; \quad \frac{\partial f}{\partial x_2} = 0$$

Como $\partial f / \partial x_1 = 1 \neq 0$, $\partial f / \partial x_2 = 2 \neq 0$ (f , función lineal, no tiene puntos críticos) no hay soluciones de (6.93) en el interior de B .

b) $g_1(x_1, x_2) = 0$; $g_2(x_1, x_2) < 0$. $g_2 = 0 \Rightarrow \lambda_2 = 0$ y (6.93) toma la forma

$$\begin{cases} \frac{\partial L}{\partial x_1} = 1 - \lambda_1 = 0 \\ \frac{\partial L}{\partial x_2} = 2 + 2\lambda_1 = 0 \\ x_1 - 2x_2 - 2 = 0; \quad \lambda_1 \leq 0 \\ x_1^2 + x_2^2 - 4 < 0; \quad \lambda_2 = 0 \end{cases}$$

Las dos primeras ecuaciones son incompatibles ($\lambda_1 = 1$ y $\lambda_1 = -1$) luego no hay ninguna solución posible en esta situación b).

c) $g_1(x_1, x_2) < 0$ ($\Rightarrow \lambda_1 = 0$); $g_2(x_1, x_2) = 0$. (6.93) toma la forma

$$\begin{cases} \frac{\partial L}{\partial x_1} = 1 - 2\lambda_2 x_1 = 0 \\ \frac{\partial L}{\partial x_2} = 2 - 2\lambda_2 x_2 = 0 \\ x_1 - 2x_2 - 2 < 0; \lambda_1 = 0 \\ x_1^2 + x_2^2 = 4; \lambda_2 \leq 0 \end{cases}$$

De las dos primeras ecuaciones vemos que ha de ser $\lambda_2 \neq 0$ y $x_2 = 2x_1$. Utilizando esto último en $x_1^2 + x_2^2 = 4$, se obtienen las posibles soluciones $P_1(2/\sqrt{5}, 4/\sqrt{5})$ y $P_2(-2/\sqrt{5}, -4/\sqrt{5})$. El punto P_2 no verifica la restricción $g_1(x_1, x_2) < 0$ ya que $g_1(-2/\sqrt{5}, -4/\sqrt{5}) = 6/\sqrt{5} - 2 \cong 0, 6833 > 0$. P_1 sí la verifica, pero, para él, $\lambda_2 = \sqrt{5}/4 > 0$. Así, pues, tampoco en esta situación hay candidatos a solución. (Obsérvese que P_1 es candidato a máximo y, de hecho, f alcanza en él el valor máximo en B .)

d) $g_1(x_1, x_2) = 0$; $g_2(x_1, x_2) = 0$, o sea, las dos restricciones se saturan. (6.93) queda

$$\begin{cases} \frac{\partial L}{\partial x_1} = 1 - \lambda_1 - 2\lambda_2 x_1 = 0 \\ \frac{\partial L}{\partial x_2} = 2 + 2\lambda_1 - 2\lambda_2 x_2 = 0 \\ x_1 - 2x_2 - 2 = 0; \lambda_1 \leq 0 \\ x_1^2 + x_2^2 - 4 = 0; \lambda_2 \leq 0 \end{cases}$$

Comenzamos resolviendo el sistema algebraico:

$$\begin{cases} x_1 - 2x_2 - 2 = 0 \\ x_1^2 + x_2^2 - 4 = 0 \end{cases}$$

sustituyendo la primera ecuación en la segunda. Se obtienen las soluciones $P_3(2, 0)$ y $P_4(-6/5, -8/5)$ (son los puntos de intersección de la recta $x_1 - 2x_2 - 2 = 0$ con la circunferencia $x_1^2 + x_2^2 - 4 = 0$). Para P_3 se tiene

$$\begin{cases} 1 - \lambda_1 - 4\lambda_2 = 0 \\ 2 + 2\lambda_1 = 0 \end{cases}$$

de donde resulta $\lambda_1 = -1$ y $\lambda_2 = 1/2$; como λ_2 es positivo, hay que descartar P_3 como posible solución. Para P_4 se tiene

$$\begin{cases} 1 - \lambda_1 + \frac{12}{5}\lambda_2 = 0 \\ 2 + 2\lambda_1 + \frac{16}{5}\lambda_2 = 0 \end{cases}$$

de donde resulta $\lambda_1 = -1/5$ y $\lambda_2 = -1/2$; como ambos son negativos, P_4 sí es candidato a solución y, de hecho, por ser el único, es la solución ya que, como decíamos, el problema tiene solución al ser aplicable el teorema de Weierstrass (nótese, además, que los vectores

$$\begin{aligned} \nabla g_1(x_1, x_2) &= (1, -2) \\ \nabla g_2(x_1, x_2) &= (2x_1, 2x_2) \end{aligned}$$

evaluados en P_4 :

$$\begin{aligned} \nabla g_1(-6/5, -8/5) &= (1, -2) \\ \nabla g_2(-6/5, -8/5) &= (-12/5, -16/5) \end{aligned}$$

son linealmente independientes).

La resolución del sistema de condiciones de Kuhn-Tucker puede resultar, salvo en problemas de tipo didáctico como los anteriores, muy complicada. Nótese que con sólo 3 restricciones ya habría que considerar $2^3 = 8$ situaciones distintas.

En relación al teorema de Kuhn-Tucker, cabe hacer **observaciones** análogas a las realizadas en el apartado precedente sobre los problemas con restricciones de igualdad:

1. La hipótesis de independencia lineal de los gradientes de las restricciones que se saturan en el óptimo es una **condición de regularidad** que se puede debilitar algo (si bien su expresión se complica respecto a la dada aquí) pero de la que no se puede prescindir totalmente. Nos sirve para ilustrarlo el mismo ejemplo 11 del apartado anterior formulado así:

$$\begin{aligned} \min \{x^2 + y^2\}, \quad (x, y) \in \mathbb{R}^2 \\ \text{s.a. } (x+1)^3 + y^2 \leq 0 \end{aligned}$$

(compruébense los detalles). Obsérvese que en la condición de regularidad está implícito el que $r \leq n$, ya que en \mathbb{R}^n no puede haber más de n vectores linealmente independientes; en otras palabras, el número de restricciones que se saturan en el óptimo no ha de ser mayor que el de variables de decisión, lo que se comprende fácilmente a la luz de los comentarios que realizamos en el apartado anterior sobre el número m de restricciones en un problema con restricciones de igualdad.

2. Como en el caso del teorema de Lagrange, las condiciones (6.90) de Kuhn-Tucker son necesarias pero no suficientes. Por ejemplo, si $f(x) = x^3$ y $g(x) = x$ y nos planteamos el problema

$$\begin{aligned} \min f(x), \quad x \in \mathbb{R} \\ \text{s.a. } g(x) \leq 0 \end{aligned} \quad (6.94)$$

veremos, por un lado, que $g'(x) = 1$ para todo x , es decir, que se satisface la condición de regularidad en todo x , y, por otro, que los valores $x^* = 0$ (que pertenece al conjunto factible y el λ^* se satura la restricción) y $\lambda^* = 0$ verifican las condiciones de Kuhn-Tucker

$$f'(x^*) = \lambda^* g'(x^*)$$

Si éstas fuesen suficientes, $x^* = 0$ sería solución de (6.94), lo que claramente no es el caso (de hecho, $x^* = 0$ es el máximo de f en $B = (-\infty, 0]$).

3. Hay que tener presente también que, en los argumentos precedentes, S , el conjunto de partida en la optimización, se supone *abierto*. Si no lo es, habrá que considerar también la evaluación de f en los posibles puntos de la frontera de S que satisfacen las restricciones, pues las condiciones necesarias (6.90) son para óptimos x^* que sean puntos interiores de S .

Problemas convexos

Supongamos que en el problema

$$\begin{aligned} \min f(x), \quad x \in S \subseteq \text{Dom } f \\ \text{s.a. } g_j(x) \leq 0, \quad j = 1, \dots, m \end{aligned} \quad (6.95)$$

el conjunto S es convexo y que las funciones f, g_j son todas convexas en S . El conjunto factible

$$B = \{x \in S; g_j(x) \leq 0, j = 1, \dots, m\}$$

es convexo por ser intersección de conjuntos convexos:

$$B = \{x \in S; g_1(x) \leq 0\} \cap \dots \cap \{x \in S; g_m(x) \leq 0\}$$

Supongamos que en $x^* \in B$ se satisfacen las condiciones de Kuhn-Tucker para el problema (6.95) y sean $(\lambda_1^*, \dots, \lambda_m^*) = \lambda^*$ sus multiplicadores asociados. Esto implica que x^* es un punto crítico de la función

$$F(x) \stackrel{\text{def}}{=} L(x, \lambda^*) = f(x) - \sum_{j=1}^m \lambda_j^* g_j(x)$$

Pero, como $\lambda_j^* \leq 0$, la función $F(x)$ es convexa en S (por ser una combinación lineal con coeficientes ≥ 0 de funciones convexas) con lo que, por los resultados sobre convexidad del apartado 6.1, se tendrá que x^* es un mínimo global de $F(x)$ en S :

$$F(x) \geq F(x^*) \text{ para todo } x \in S$$

es decir

$$f(x) - \sum_{j=1}^m \lambda_j^* g_j(x) \geq f(x^*) - \sum_{j=1}^m \lambda_j^* g_j(x^*), \quad x \in S \quad (6.96)$$

Restringiéndonos al conjunto factible

$$B = \{x \in S; g_j(x) \leq 0, j = 1, \dots, m\}$$

y teniendo en cuenta que $\lambda_j^* g_j(x^*) = 0$ por (6.90), (6.96) implica que

$$f(x) \geq f(x^*) \text{ para todo } x \in B$$

es decir, que x^* es solución (global) del problema (6.95). Vemos, pues, que, bajo hipótesis extra de convexidad, las condiciones necesarias de Kuhn-Tucker pasan a ser también condiciones suficientes para la existencia de mínimo. (En el ejemplo 18 se dan estas hipótesis de convexidad y, en consecuencia, se puede afirmar que el punto P_4 es solución del problema sin necesidad de apelar al teorema de Weierstrass.). Fórmúlese como ejercicio el resultado análogo para un problema de maximización. Compruébese que un programa lineal del tipo (6.95), o sea, uno en el que tanto la función objetivo como las funciones que definen las restricciones son lineales, es convexo, y, por tanto, todo punto en el que se verifiquen las condiciones de Kuhn-Tucker —fórmúlese como ejercicio— es de hecho óptimo global del problema.

Condiciones suficientes de optimalidad

Son condiciones que se aplican a los puntos que satisfacen las condiciones necesarias de Kuhn-Tucker. En consecuencia, si no se satura ninguna restricción, se trata de un punto crítico interior de la función objetivo y las condiciones suficientes son las de óptimos sin restricciones del apartado 6.1, y, si se satura en el punto bajo estudio alguna de las restricciones, las condiciones suficientes serán las de los problemas con restricciones de igualdad relativas a las restricciones saturadas.

Significado de los multiplicadores de Kuhn-Tucker

Aquí también se puede dar una interpretación de los multiplicadores como medida de la sensibilidad del valor óptimo a cambios pequeños en el nivel de cada restricción $g(x) \leq c$ que se deriva de la que hemos comentado en los problemas con restricciones de igualdad (¿qué significaría, en particular, que $\lambda_j = 0$ para una restricción que no se satura: $g(x) < c$? ¿y el signo negativo, en un problema de minimización, para una que se satura?).

El problema general de optimización

El problema general de optimización es aquel que incluye a la vez restricciones de igualdad y restricciones de desigualdad:

$$\begin{array}{ll} \text{opt } f(x_1, \dots, x_n), & (x_1, \dots, x_n) \in S \subseteq \text{Dom } f \\ \text{s.a.} & \\ \left\{ \begin{array}{l} g_1(x_1, \dots, x_n) \leq 0 \\ g_2(x_1, \dots, x_n) \leq 0 \\ \vdots \\ g_m(x_1, \dots, x_n) \leq 0 \end{array} \right. & \left\{ \begin{array}{l} h_1(x_1, \dots, x_n) = 0 \\ h_2(x_1, \dots, x_n) = 0 \\ \vdots \\ h_s(x_1, \dots, x_n) = 0 \end{array} \right. \end{array} \quad (6.97)$$

(las restricciones de igualdad han de ser en número menor que el de variables de decisión: $s < n$). De los teoremas de Lagrange y Kuhn-Tucker se deduce como corolario el siguiente

Teorema 6.18 Sean $f, g_1, \dots, g_m, h_1, \dots, h_s$, $s < n$, funciones de clase C^1 en un subconjunto abierto $S \subseteq \mathbb{R}^n$ con valores en \mathbb{R} . Supongamos que $x^* = (x_1^*, \dots, x_n^*)$ es un óptimo local de f en el conjunto

$$B = \{x \in S; g_j(x) \leq 0, j = 1, \dots, m; h_k(x) = 0, k = 1, \dots, s\}$$

o sea, una solución local del problema (6.97). Reordenando las funciones g_j si es necesario, podemos suponer que las restricciones de desigualdad que se saturan en x^* son $g_1(x^*) = 0, \dots, g_r(x^*) = 0$, con $r \leq m$. Pues bien, si los vectores $\{\nabla g_1(x^*), \dots, \nabla g_r(x^*), \nabla h_1(x^*), \dots, \nabla h_s(x^*)\}$ son linealmente independientes, entonces existen constantes $\lambda_1^*, \dots, \lambda_m^*$ y μ_1^*, \dots, μ_s^* tales que

$$\left\{ \begin{array}{l} \nabla f(x^*) = \lambda_1^* \nabla g_1(x^*) + \dots + \lambda_m^* \nabla g_m(x^*) + \mu_1^* \nabla h_1(x^*) + \dots + \mu_s^* \nabla h_s(x^*) \\ y, \text{ para todo } j = 1, \dots, m, \quad \lambda_j^* g_j(x^*) = 0 \text{ y} \\ \lambda_j^* \leq 0 \text{ si } x^* \text{ es un mínimo o } \lambda_j^* \geq 0 \text{ si } x^* \text{ es un máximo.} \end{array} \right. \quad (6.98)$$

En la práctica, el esquema de trabajo para intentar resolver el problema (6.97) es el mismo que antes:

- (i) Se forma el lagrangiano del problema:

$$L(x, \lambda, \mu) = f(x) - \lambda_1 g_1(x) - \dots - \lambda_m g_m(x) - \mu_1 h_1(x) - \dots - \mu_s h_s(x)$$

- (ii) Se resuelve el sistema de condiciones de Kuhn-Tucker:

$$\begin{cases} \frac{\partial L}{\partial x_i} = \frac{\partial f}{\partial x_i} - \sum_{j=1}^m \lambda_j \frac{\partial g_j}{\partial x_i} - \sum_{k=1}^s \mu_k \frac{\partial h_k}{\partial x_i} = 0, \quad i = 1, \dots, n \\ g_j(x) \leq 0, \quad \lambda_j \leq 0 \text{ (o bien } \lambda_j \geq 0), \quad \lambda_j g_j(x) = 0, \quad j = 1, \dots, m \\ h_k(x) = 0, \quad k = 1, \dots, s \end{cases} \quad (6.99)$$

y sus soluciones $(x_1^*, \dots, x_n^*, \lambda_1^*, \dots, \lambda_m^*, \mu_1^*, \dots, \mu_s^*)$ (en las que, recuérdese, los *multiplicadores de Kuhn-Tucker* asociados a las restricciones de desigualdad han de ser ≤ 0 para un problema de minimización y ≥ 0 para uno de maximización, mientras que los multiplicadores asociados a las restricciones de igualdad pueden tener cualquier signo) proporcionan, quedándose con las n primeras coordenadas (x_1^*, \dots, x_n^*) , *candidatos* a soluciones del problema (6.97). En cuanto a la resolución efectiva del problema cabe hacer las mismas observaciones que anteriormente.

6.4 Problemas

1. Hallar y clasificar los puntos críticos de las funciones:

i) $f(x, y) = x^4 + 8x^2 + y^2 - 4y$

ii) $f(x, y) = -x^2 + xy - y^2 + 2x - y + 5$

iii) $f(x, y) = (x^2 - 2x) \cdot e^y$

iv) $f(x, y) = e^{(x-y)^2 + x^2}$

2. Analizando la concavidad/convexidad de las funciones del ejercicio anterior, determinar si los óptimos locales obtenidos son globales en \mathbb{R}^2 .
3. (a) Demostrar que la intersección de conjuntos convexos siempre da como resultado otro conjunto convexo. ¿Qué se puede decir, en general, de la unión de conjuntos convexos?

(b) Analizar la convexidad de los siguientes conjuntos:

$$S_1 = \{(x, y) \in \mathbb{R}^2; (x-1)^2 + (y-1)^2 \leq 1, y \leq x\}$$

$$S_2 = \{(x, y) \in \mathbb{R}^2; x^2 - y \leq 1, x + y \geq 1\}$$

$$S_3 = \{(x, y) \in \mathbb{R}^2; xy \geq 1, y \leq ax\} \text{ (con } a > 0\text{)}$$

$$S_4 = \{(x, y) \in \mathbb{R}^2; xy \geq 1, x \geq 0, y \leq ax\} \text{ (con } a > 0\text{)}$$

$$S_5 = \{(x, y) \in \mathbb{R}^2; 4x^2 + y^2 \leq 4, x^2 + 4y^2 \leq 4\}$$

$$S_6 = \{(x, y) \in \mathbb{R}^2; x^2 - y^2 \geq 1, (x-1)^2 + y^2 \leq 1\}$$

$$S_7 = \{(x, y) \in \mathbb{R}^2; 4x - y^2 \geq 0, y - x^3 \geq 0\}$$

$$S_8 = \{(x, y) \in \mathbb{R}^2; x \geq 1, \ln x \leq y \leq e^x\}$$

$$S_9 = \{(x, y, z) \in \mathbb{R}^3; x^2 + 9y^2 + z^2 \leq 1, 5x + y - 2z \leq 0\}$$

$$S_{10} = \{(x, y, z) \in \mathbb{R}^3; x^2 + y^2 \geq 1, x^2 + y^2 - z^2 \leq 0\}$$

4. Demostrar que una función $f: \mathbb{R}^n \rightarrow \mathbb{R}$ es convexa en \mathbb{R}^n si, y sólo si, el conjunto (denominado *epigrafo*)

$$\{(x, y) \in \mathbb{R}^{n+1}; y \geq f(x)\}$$

es un conjunto convexo. Establecer una caracterización similar para funciones cóncavas.

5. (a) Sea f una función tal que $f(x) \geq 0$ para todo $x \in \mathbb{R}^n$. Demostrar que, si el problema

$$\min \sqrt{f(x)}$$

tiene solución global y f es diferenciable en \mathbb{R}^n , entonces ésta se alcanza en un punto crítico de f .

- (b) Utilizando lo anterior, determinar la mínima distancia existente entre un punto de la parábola $y = x^2$ y un punto de la recta $y = 2x - 6$.

6. Comprobar que $(0, 0)$ es punto crítico de las funciones

$$i) f(x, y) = x^4(x^2 + y^2 - 1) \quad ii) f(x, y) = x^4 - y^2$$

Analizando el signo de estas funciones, establecer de qué tipo de punto crítico se trata.

7. Dadas las funciones:

$$i) f(x, y) = y^2 - 4xy + 4x^2$$

$$ii) f(x, y, z) = -x^2 + 4xy + 2xz - 6yz - 2z^2$$

se pide:

- Determinar sus puntos críticos.
 - Analizar la concavidad/convexidad de estas funciones.
 - Utilizando lo obtenido en (b), clasificar los puntos críticos hallados en (a).
8. Determinar y clasificar los puntos críticos de la función

$$f(x, y) = \frac{4x}{x^2 + y^2 + 1}$$

Demostrar que los óptimos locales obtenidos son globales en \mathbb{R}^2 .

9. Comprobar que la función

$$f(x, y) = x^2(1 + y)^3 + y^2$$

tiene un único punto crítico que es un mínimo local. ¿Es global? Comparar con lo que ocurre en una situación semejante para funciones de una variable.

10. Un monopolista vende en un mercado un bien que puede producir en dos plantas distintas. Si las funciones de coste en sus dos plantas están dadas por

$$C_1(q_1) = 2q_1^2 + 4, \quad C_2(q_2) = 6q_2^2 + 8,$$

siendo q_1 y q_2 las cantidades respectivamente producidas en cada una de las dos plantas, y la función de demanda para su producto es $p = 88 - 4q$, donde p es el precio de venta del bien y $q = q_1 + q_2$ la cantidad demandada (que se supone coincide con la ofertada), calcular la cantidad que ha de producir en cada planta para maximizar los beneficios. Para este nivel de producción, determinar el precio de venta del producto y el beneficio obtenido.

11. Hallar los valores de a y b para que la función

$$f(x, y) = ax^3 + 3bxy^2 - 15a^2x - 12y + 5$$

tenga un mínimo local en el punto $(2, 1)$.

12. Una empresa fabrica un bien a partir de dos materias primas de acuerdo con la función de producción

$$q(x, y) = \ln x + \ln y,$$

siendo x e y las cantidades utilizadas de cada materia prima y $q(x, y)$ la cantidad obtenida de producto final. Sabiendo que la función de coste viene dada por

$$C(x, y) = 3x + 2y + 5$$

y que el precio de venta del bien es $p = 6$, determinar el nivel de producción que maximiza los beneficios.

13. Dada la función

$$F(x, y, z) = z - (x^2 + y^2 + 1) \cdot e^z$$

se pide:

- Determinar y clasificar sus puntos críticos.
- Hacer un estudio de la concavidad/convexidad de $F(x, y, z)$ e intentar aplicarlo, junto con lo obtenido en (a), a la detección de conjuntos $D \subseteq \mathbb{R}^3$ en los que se pueda afirmar la existencia de óptimos globales de f .
- Estudiar si la ecuación

$$F(x, y, z) = k$$

para un k apropiado (¿cuál?) define en un entorno del punto $(0, 0, -1)$ una función $z = f(x, y)$.

- Comprobar que $f(x, y)$ tiene en $(0, 0)$ un punto crítico y establecer de qué tipo es.
- Escribir la ecuación del plano tangente a $z = f(x, y)$ en el punto $(0, 0, -1)$ y el polinomio de Taylor de segundo grado de $f(x, y)$ en el entorno de $(0, 0)$.

(f) Resolver el problema

$$\text{opt } F(x, y, z)$$

$$\text{s.a.}$$

$$x^2 + y^2 = 1$$

$$x + z = 1$$

14. Resolver, estudiando la globalidad, los siguientes problemas con restricciones de igualdad:

$$i) \begin{cases} \text{opt } f(x, y, z) = 3 + x^2 + 2y^2 + 4y - 2x + (z - 2)^2 \\ \text{s.a.} \\ 2x + 4y + z = 0 \end{cases}$$

$$ii) \begin{cases} \text{opt } f(x, y) = x^2 y \\ \text{s.a.} \\ x^2 + y^2 = 1 \end{cases}$$

$$iii) \begin{cases} \text{opt } f(x, y, z) = z \\ \text{s.a.} \\ x^2 + y^2 = 4 \\ x + y + z = 5 \end{cases}$$

$$iv) \begin{cases} \text{opt } f(x, y) = x^4 + 3y^4 + 1 \\ \text{s.a.} \\ x - y = 0 \end{cases}$$

$$v) \begin{cases} \text{opt } f(x, y, z) = -x^4 - 2y^2 - z^2 + 20 \\ \text{s.a.} \\ y = x \\ 2x - y - z = 5 \end{cases}$$

15. Descomponer el número 9 en suma de tres sumandos de manera que la suma de los cuadrados de éstos sea mínima.

16. (a) Demostrar mediante el método de los multiplicadores de Lagrange que dado un hiperplano en \mathbb{R}^n

$$a_1 x_1 + \dots + a_n x_n = d$$

existe en él un punto (x_1, \dots, x_n) cuya distancia al origen es menor que la distancia del origen a cualquier otro punto del hiperplano.

- (b) Resolver el problema

$$\min\{x^2 + y^2 + z^2\}$$

$$\text{s.a.} \begin{cases} x + 2y + z = 1 \\ 2x - y - 3z = 4 \end{cases}$$

Dar una interpretación geométrica del resultado obtenido. Estudiar también el correspondiente problema de maximización.

17. Analizar la posible aplicación del teorema de Lagrange a la resolución del problema

$$\min\{y\} \text{ s.a. } y^3 - x^2 = 0$$

18. Un empresario ha de pagar dos impuestos: el de sociedades y el de la renta de las personas físicas. El de sociedades es de tipo fijo $t_S = 0.3$ (en tanto por uno) y afecta a los beneficios empresariales. En cambio, el impuesto de la renta es progresivo; su tipo impositivo depende de los ingresos, en concreto $t_R = 0.1 \cdot I$, siendo I el ingreso (sueldo) obtenido. Supongamos que el empresario puede fijar su sueldo, distribuyendo entre éste y los beneficios declarados de la empresa, los beneficios verdaderamente obtenidos. Si los beneficios que obtiene la empresa son de 3 unidades monetarias, ¿qué sueldo se fijará para minimizar el pago de impuestos?
19. Un monopolista vende dos productos A y B cuyas funciones de demanda vienen definidas por las ecuaciones

$$\begin{aligned} p_1 - 12 + 2x_1 &= 0 \\ p_2 - 32 + 4x_2 &= 0 \end{aligned}$$

en donde p_1, p_2 y x_1, x_2 son, respectivamente, los precios y las cantidades producidas (y vendidas) de los productos A y B . Sabiendo que su función de coste es

$$C(x_1, x_2) = x_1^2 + 2x_1x_2 + x_2^2$$

determinar la producción que maximiza el beneficio.

20. Una empresa que fabrica tres bienes (A , B y C) sabe que sus costes diarios (medidos en miles de pesetas) vienen dados por la función

$$C(x, y, z) = \frac{1}{2}x^4 + x^2 + y^2 + y + 3z^2 + z$$

donde las variables x, y, z denotan las cantidades (en toneladas) diarias producidas respectivamente de los bienes A, B y C. Para fabricar estos bienes se utilizan dos máquinas: la primera está en funcionamiento 10 horas y la segunda 12 horas cada día laborable. Las horas necesarias en cada máquina para elaborar una tonelada de cada bien se indican en la siguiente tabla

	Máquina 1	Máquina 2
Bien A	1	2
Bien B	1	1
Bien C	1	1

Se pide:

- Formular el programa con restricciones de igualdad que permite determinar la producción diaria de cada bien que maximiza los beneficios de la empresa.
 - Resolver el programa anterior aplicando las condiciones de primer y segundo orden de Lagrange y estudiando la globalidad.
 - Si la empresa tiene la posibilidad de aumentar ligeramente el tiempo de funcionamiento de una de las dos máquinas, ¿qué máquina preferirá?. Justifíquese la respuesta.
- Enunciar el teorema de la envolvente (dado en la nota 2 que sigue al teorema 6.15) para el caso particular de un problema con 2 variables, 1 restricción y dependiente de un parámetro $c \in \mathbb{R}$. Demuéstrese este resultado supuesto que el problema tiene solución para todo $c \in \mathbb{R}$ y que la función de valor óptimo es de clase C^1 .
 - Un inversor dispone de 4 unidades monetarias que quiere invertir en tres títulos mobiliarios. Se denomina cartera de valores al vector $x = (x_1, x_2, x_3)$, donde x_i , $i = 1, 2, 3$, denota la cantidad de dinero invertida en el título i . Sean $\mu_1 = 0.6$, $\mu_2 = 0.3$ y $\mu_3 = 0.5$ las rentabilidades anuales esperadas (en tanto por uno) de cada uno de ellos. El riesgo en que se incurre al realizar la inversión viene dado por la función

$$R(x) = x_1^2 + 2x_2^2 + 2x_3^2$$

El inversor es conservador y su objetivo es minimizar el riesgo para una rentabilidad media prefijada $\eta = 0.5$. Se pide:

- Determinar la cartera óptima.

- (b) ¿Cómo afecta al riesgo un aumento pequeño de la inversión total?

23. Dada la función

$$f(x, y) = (x + y)^3 - 12xy$$

se pide:

- Determinar y clasificar sus puntos críticos.
- Hacer un estudio de la concavidad/convexidad de $f(x, y)$ e intentar aplicarlo, junto con lo obtenido en (a), a la detección de conjuntos $D \subseteq \mathbb{R}^2$ en los que se pueda afirmar la existencia de óptimos globales de f .
- Determinar el valor máximo y mínimo de $f(x, y)$ en el conjunto $B = \{(x, y) \in \mathbb{R}^2; x^2 + y^2 \leq 4\}$. (Indicación: Utilícese en este punto el método de los multiplicadores de Lagrange)

24. Dada la función

$$f(x, y) = xy \cdot \ln(x^2 + y^2)$$

se pide:

- Determinar y clasificar sus puntos críticos.
- ¿Son los óptimos locales de f obtenidos en (a) globales en \mathbb{R}^2 ?
- Resolver el problema

$$\text{opt } f(x, y), \quad (x, y) \in S = \{(x, y) \in \mathbb{R}^2; x^2 + y^2 \leq 9\}$$

25. Dado el programa

$$\begin{array}{ll} \max & f(x, y) = y \\ \text{s.a.} & \\ & x^2 + y^2 \leq 6 \\ & x^2 - y \geq 0 \end{array}$$

se pide:

- Resolverlo gráficamente.
- Comprobar que en los puntos obtenidos en (a) se verifican las condiciones de Kuhn-Tucker.
- ¿Se verifican las condiciones de Kuhn-Tucker en el punto $(0, 0)$?, ¿es solución del programa?. Explicar razonadamente lo obtenido.

26. Resolver los siguientes problemas:

$$i) \begin{cases} \max f(x, y) = 2x^2 - 3y^2 - 2x \\ \text{s.a.} \\ x^2 + y^2 \leq 1 \end{cases}$$

$$ii) \begin{cases} \min f(x, y, z) = x^2 + y^2 + z^2 - 4y - 2z \\ \text{s.a.} \\ y + z \geq 1 \\ x^2 - z \leq 5 \end{cases}$$

$$iii) \begin{cases} \max f(x, y) = x^2 - 2y \\ \text{s.a.} \\ x + y \geq 4 \\ x^2 + y^2 \leq 16 \end{cases}$$

$$iv) \begin{cases} \min f(x, y) = x^2 + (y - 2)^2 \\ \text{s.a.} \\ x + y \leq 1 \\ y - x \leq 1 \\ y \geq 0 \end{cases}$$

27. Analizar la posible aplicación del teorema de Kuhn-Tucker a la resolución del problema

$$\begin{cases} \min f(x, y) = (x + 3)^2 + y^2 \\ \text{s.a.} \\ y - x^3 \leq 0 \\ y + x^3 \geq 0 \end{cases}$$

28. Una empresa de automóviles ha firmado un contrato para entregar 150 coches eléctricos durante los 3 próximos meses, comprometiéndose a entregar, al menos, 50 coches cada mes. El coste de producir s coches en un mes es de s^2 unidades monetarias y el coste de almacenar un coche producido en un mes hasta entregarlo al mes siguiente es de 20 u. m.. Se pide:

- Determinar la política óptima de producción mensual minimizando los costes totales de producción y almacenamiento.
- ¿Por cuánto dinero, como mínimo, la empresa entregaría un coche adicional más el último mes?
- ¿Por cuánto dinero, como mínimo, la empresa entregaría un coche adicional más al mes durante cada uno de los tres meses?

Parte II

Funciones Vectoriales

Capítulo 7

Topología del espacio R^n

En éste y los dos capítulos que siguen vamos a generalizar a funciones vectoriales buena parte de los conceptos y resultados que hemos visto hasta aquí así como a realizar las demostraciones pendientes de resultados fundamentales como el teorema de Weierstrass sobre máximos y mínimos y el teorema de la función implícita.

Recordemos que por **función vectorial** entendemos una función f cuyo dominio es un subconjunto de R^n y cuya imagen o recorrido es un subconjunto de R^m con $m > 1$, o sea, una correspondencia que asigna a cada $x = (x_1, \dots, x_n)$ de un subconjunto de R^n , que denominamos *dominio* de f y denotamos por $Dom f$, un punto (vector) $y = f(x)$ de R^m bien determinado. Escribiremos $f: D \rightarrow R^m$ para indicar que estamos considerando la "actuación" de f en un subconjunto $D \subseteq Dom f$. (Véase el apartado 1.1 para las definiciones generales relativas a funciones.)

Dada $f: Dom f \rightarrow R^m$, las coordenadas (y_1, \dots, y_m) de $y = f(x)$ están unívocamente determinadas por $x \in Dom f$; en otras palabras, dada una función $f: Dom f \rightarrow R^m$ están bien definidas las funciones escalares $f_i: Dom f \rightarrow R$, $i = 1, \dots, m$, mediante

$$f_i(x) = y_i$$

Estas funciones se llaman **funciones coordenadas o componentes** de f ; escribiremos $f(x) = (f_1(x), \dots, f_m(x))$ o, simplemente, $f = (f_1, \dots, f_m)$. A través de las componentes, podremos extender fácilmente a funciones vectoriales muchos de los conceptos y resultados vistos para funciones escalares.

Para llevar a cabo esta tarea, comenzamos en el apartado 7.1 repitiendo en un estilo algo más abstracto y sistemático cuestiones ya tratadas en el capítulo 2, con el objetivo de disponer de un resumen conciso de las bases que permiten tratar con precisión la idea de *proximidad* y, con ella, introducir los

conceptos fundamentales de *límite* y *continuidad*. Aquí, como en el resto de apartados, no nos detendremos en la *motivación* de las ideas en juego cuando ya ha sido dada en el lugar correspondiente de la primera parte del libro. Las demostraciones se llevan al apartado de problemas como "ejercicios de abstracción".

El apartado 7.2 trata de las sucesiones de puntos de \mathbb{R}^n y de su convergencia; las definiciones y resultados que se exponen derivan directamente de los correspondientes para sucesiones de números reales, que suponemos conocidos por el lector del curso de funciones de una variable. Como para éstas, se podrán caracterizar mediante sucesiones las nociones de límite y continuidad así como distintos tipos de conjuntos sobre los que están definidas las funciones bajo consideración.

En 7.3 se introducen unos conjuntos de gran importancia en el análisis matemático, los conjuntos **compactos**. En el apartado 7.4 establecemos ya las nociones de límite y continuidad de funciones vectoriales, generalizando y completando lo visto en el capítulo 2 para funciones escalares, y demostramos, por fin, el teorema de Weierstrass sobre máximos y mínimos.

En 7.5 veremos que para tratar las cuestiones *topológicas* en \mathbb{R}^n se puede utilizar la *norma* que nos resulte más conveniente y en 7.6 aplicaremos todo lo anterior al caso particular, pero importante, de las aplicaciones *lineales* de \mathbb{R}^n en \mathbb{R}^m , dadas por matrices.

En el apartado 7.7 profundizamos en la propiedad de **conexión** con el objetivo de alcanzar cierto grado de completación en el estudio topológico de \mathbb{R}^n , pero la noción de conjunto "conexo por poligonales" que dimos en el apartado 4.1 es, con su simplicidad, suficiente para un curso de cálculo diferencial y, como demostraremos, no supone pérdida alguna de generalidad. En 7.8 haremos algunos comentarios sobre la extensión de los conceptos introducidos y los resultados establecidos a espacios más generales que \mathbb{R}^n .

7.1 Conjuntos abiertos y cerrados

El punto de partida es la noción de *norma* de un punto $x = (x_1, \dots, x_n)$ de \mathbb{R}^n :

$$|x| = (x_1^2 + \dots + x_n^2)^{\frac{1}{2}} \quad (7.1)$$

y la correspondiente noción de distancia entre dos puntos $x = (x_1, \dots, x_n)$, $y = (y_1, \dots, y_n)$ de \mathbb{R}^n

$$d(x, y) = |x - y| = \sqrt{(x_1 - y_1)^2 + \dots + (x_n - y_n)^2} \quad (7.2)$$

Según vimos en el apartado 1.4, las propiedades fundamentales de la norma son:

$$|x| \geq 0 \text{ para todo } x \in \mathbb{R}^n \text{ y } |x| = 0 \text{ si y sólo si } x = 0 \quad (7.3)$$

$$|x + y| \leq |x| + |y| \text{ (desigualdad triangular)} \quad (7.4)$$

$$|cx| = |c| |x|, \quad c \in \mathbb{R} \quad (7.5)$$

y, deducidas de éstas, las de la distancia:

$$d(x, y) \geq 0 \text{ y } d(x, y) = 0 \text{ si y sólo si } x = y \quad (7.6)$$

$$d(x, y) = d(y, x) \quad (7.7)$$

$$d(x, z) \leq d(x, y) + d(y, z) \text{ (desigualdad triangular)} \quad (7.8)$$

Con la distancia se definen los instrumentos básicos para medir el grado de proximidad en \mathbb{R}^n :

Definición 7.1 Dados $x_0 \in \mathbb{R}^n$ y $\varepsilon > 0$, la **bola abierta** de centro x_0 y radio ε es el conjunto

$$B_\varepsilon(x_0) = \{x \in \mathbb{R}^n; |x - x_0| < \varepsilon\}$$

Un conjunto $U \subset \mathbb{R}^n$ se dice que es un **entorno** de x_0 si existe $\varepsilon > 0$ tal que $B_\varepsilon(x_0) \subset U$ (en particular, $B_\varepsilon(x_0)$ se llama también ε -entorno de x_0).

Definición 7.2 Un conjunto A se dice que es **abierto** si es entorno de todos sus puntos, o sea, si para todo $x \in A$ existe $\varepsilon > 0$ tal que $B_\varepsilon(x) \subset A$.

Por así decir, un conjunto es abierto si ocurre que cuando un punto x pertenece a él, también pertenecen todos los puntos suficientemente próximos a x . Vimos en el capítulo 2 que toda bola abierta $B_\varepsilon(x_0)$ es un conjunto abierto.

Las propiedades esenciales de los conjuntos abiertos de \mathbb{R}^n son:

Proposición 7.1 (a) \mathbb{R}^n y el conjunto vacío \emptyset son conjuntos abiertos.

(b) La unión de cualquier colección de conjuntos abiertos de \mathbb{R}^n es un conjunto abierto de \mathbb{R}^n .

- (c) La intersección de una colección finita de conjuntos abiertos de \mathbb{R}^n es un conjunto abierto de \mathbb{R}^n .

Demostración: Hágase como ejercicio (problema 2).

La familia de todos los conjuntos abiertos de \mathbb{R}^n , con las propiedades esenciales recogidas en esta proposición, constituye la **topología** de \mathbb{R}^n ; es la estructura sobre la que se asienta la noción de continuidad de una función. En la Nota 1 de final de capítulo comentaremos la extensión de este concepto a conjuntos y situaciones más generales que \mathbb{R}^n .

Las posibles "posiciones" de un punto respecto a un conjunto se clasifican de acuerdo con la siguiente

Definición 7.3 Un punto x se dice que es un **punto interior** del conjunto A si existe un $\varepsilon > 0$ tal que $B_\varepsilon(x) \subset A$. Si existe $\varepsilon > 0$ tal que $B_\varepsilon(x) \subset A^c = \mathbb{R}^n \setminus A$ (equivalentemente, $B_\varepsilon(x) \cap A = \emptyset$) se dice que x es un **punto exterior** a A . Finalmente, si toda bola $B_\varepsilon(x)$ centrada en x contiene al menos un punto de A y al menos uno de su complementario A^c , entonces se dice que x es un **punto frontera** de A (figura 4.4).

Por definición, un punto interior es necesariamente un punto de A , un punto exterior de A es un punto (interior) de A^c y un punto frontera de A —que puede pertenecer a A o a su complementario A^c — también es un punto frontera de A^c ya que $(A^c)^c = A$. Obsérvese que se trata realmente de una clasificación pues, dado un conjunto A de \mathbb{R}^n , un punto $x \in \mathbb{R}^n$ es o bien punto interior de A , o bien punto exterior o bien punto frontera.

Definición 7.4 El conjunto de puntos interiores de A se denomina **interior** de A y se representa por $\text{int } A$. El conjunto de todos los puntos frontera de A es la **frontera** de A ; se denota por ∂A o $\text{Fr } A$. El **exterior** de A , denotado por $\text{Ext } A$, es el conjunto de puntos exteriores de A . La **adherencia** o **cierre** de A es la unión de su interior y su frontera y se denota por \bar{A} : $\bar{A} = \text{int } A \cup \partial A$.

El interior, la frontera y el exterior de A son conjuntos disjuntos dos a dos y su unión es todo \mathbb{R}^n . Obsérvese que $\text{Ext } A = \text{Int } (A^c) = (\bar{A})^c$ y que, de hecho, $\bar{A} = A \cup \partial A$. Para todo conjunto A se tiene $\text{int } A \subseteq A \subseteq \bar{A}$.

Con estas nuevas definiciones, es claro que un conjunto es abierto si y sólo si todos sus puntos son interiores, o sea, $\text{int } A = A$, lo que equivale a decir que A no contiene a ninguno de sus puntos frontera. La situación totalmente opuesta lleva a la siguiente

Definición 7.5 Un conjunto A se dice que es **cerrado** si contiene a todos sus puntos frontera, o sea, si $A \supset \partial A$ (equivalentemente, si $A = \bar{A}$).

Como vimos en el capítulo 2, hay conjuntos que no son ni abiertos ni cerrados (contienen a algunos de sus puntos frontera pero no a todos).

La caracterización siguiente permite una definición alternativa de conjunto cerrado:

Proposición 7.2 Un conjunto es cerrado si y sólo si su complementario es abierto.

Demostración: Ejercicio (problema 3).

EJEMPLO. El conjunto

$$\bar{B}_\varepsilon(x_0) = \{x \in \mathbb{R}^n; |x - x_0| \leq \varepsilon\}$$

se llama **bola cerrada** de centro x_0 y radio ε . Es, en efecto, un conjunto cerrado y, en coherencia con su denotación, es la adherencia de la bola abierta $B_\varepsilon(x_0)$ (y se tiene $\partial B_\varepsilon(x_0) = \partial \bar{B}_\varepsilon(x_0) = \{x \in \mathbb{R}^n; |x - x_0| = \varepsilon\} \stackrel{\text{def}}{=} S_\varepsilon(x_0)$; problema 4).

De las proposiciones 7.1 y 7.2 se derivan las siguientes propiedades de los conjuntos cerrados:

Proposición 7.3 1.

- (a) \mathbb{R}^n y el conjunto vacío \emptyset son conjuntos cerrados.
- (b) La intersección de cualquier colección de conjuntos cerrados de \mathbb{R}^n es un conjunto cerrado de \mathbb{R}^n .
- (c) La unión de una colección finita de conjuntos cerrados de \mathbb{R}^n es un conjunto cerrado de \mathbb{R}^n .

Demostración: Hágase como ejercicio (problema 5).

Conviene profundizar en la estructura de la adherencia de un conjunto: un punto x de \bar{A} puede o no pertenecer a A pero en cualquiera de los dos casos cumple (sea de $\text{int } A$ o de ∂A) la condición de que en todo ε -entorno suyo existe algún punto de A , es decir

$$\{x \in \bar{A}\} \iff \{B_\varepsilon(x) \cap A \neq \emptyset \text{ para todo } \varepsilon > 0\} \quad (7.9)$$

(claramente, se puede sustituir "todo ε -entorno" por "todo entorno").

Si $x \in \bar{A}$ es un punto interior de A , entonces, por definición, todos los puntos suficientemente próximos a x son también de A ; si se trata de un punto frontera, puede ocurrir que x , aun perteneciendo a A , esté aislado del resto de puntos de A o que, por el contrario, en todo entorno suyo existan puntos de A diferentes de él mismo; esto lleva a la siguiente

Definición 7.6 Un punto $x \in A$ se dice que es un **punto aislado** de A si existe $\varepsilon > 0$ tal que $B_\varepsilon(x) \cap A = \{x\}$. Un punto $x \in \mathbb{R}^n$ se dice que es un **punto de acumulación** del conjunto A si para todo $\varepsilon > 0$ existe $y \in B_\varepsilon(x) \cap A$ con $y \neq x$.

(Es claro que en esta definición se puede sustituir “existe $\varepsilon > 0$ tal que $B_\varepsilon(x) \cap A = \{x\}$ ” por “existe un entorno U de x tal que $U \cap A = \{x\}$ ” y “para todo $\varepsilon > 0$ existe $y \in B_\varepsilon(x) \cap A$ con $y \neq x$ ” por “para todo entorno U de x existe $y \in U \cap A$ con $y \neq x$ ”).

Un punto interior de A es siempre punto de acumulación de A , mientras que un punto frontera es o bien punto aislado (y, entonces, por definición, pertenece a A) o bien punto de acumulación (y, en este caso, puede o no pertenecer a A). Se tiene entonces que

$$\bar{A} = A \cup \{x \in \mathbb{R}^n; x \text{ es punto de acumulación de } A\} \quad (7.10)$$

y, como consecuencia, que

$$\boxed{A \text{ es cerrado si y sólo si contiene a sus puntos de acumulación}} \quad (7.11)$$

Por otra parte, si llamamos **punto adherente** o **punto de adherencia** a un punto de \bar{A} se tiene que un punto de adherencia es o bien aislado o bien de acumulación.

7.2 Convergencia en \mathbb{R}^n

La noción de convergencia de una sucesión de puntos de \mathbb{R}^n es la generalización natural de la correspondiente noción en la recta real \mathbb{R} , utilizando la norma de \mathbb{R}^n en lugar del valor absoluto de números reales:

Definición 7.7 Una sucesión $\{x_k\} = \{(x_1^k, \dots, x_n^k)\}$, $k \in \mathbb{N}$, de puntos de \mathbb{R}^n se dice que **converge** al punto $x = (x_1, \dots, x_n)$ si para todo $\varepsilon > 0$ existe un entero positivo $N = N(\varepsilon)$ tal que $|x_k - x| < \varepsilon$ para todo $k \geq N$; el punto x se dice que es el **límite** de la sucesión $\{x_k\}$ y, como en \mathbb{R} , se escribe

$$\lim_{k \rightarrow \infty} x_k = x \quad \text{o} \quad x_k \xrightarrow{k \rightarrow \infty} x$$

Observaciones:

1. $\lim_{k \rightarrow \infty} x_k = x$ en \mathbb{R}^n si y sólo si se tiene la convergencia en \mathbb{R}

$$\lim_{k \rightarrow \infty} |x_k - x| = 0$$

(y nótese también que $x_k \xrightarrow{k \rightarrow \infty} x$ implica $|x_k| \xrightarrow{k \rightarrow \infty} |x|$).

2. La convergencia $\lim_{k \rightarrow \infty} x_k = x$ se suele expresar de manera coloquial diciendo que cada ε -entorno de x contiene a todos los términos x_k con k suficientemente grande (o, también, a todos los términos de la sucesión salvo un número finito de ellos).
3. La referencia " $k \rightarrow \infty$ " se suele omitir cuando se deduce claramente del contexto.

Teorema 7.1 $x_k \rightarrow x$ si y sólo si $x_i^k \rightarrow x_i$ para cada $i = 1, \dots, n$.

Demostración: Hágase como ejercicio teniendo en cuenta que para cualquier punto $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ se tiene

$$\max |x_i| \leq |x| \leq \sqrt{n} \cdot \max |x_i|$$

Este resultado nos dice, en definitiva, que una sucesión de puntos de \mathbb{R}^n es convergente si se tiene convergencia coordinada a coordinada, lo que permite transcribir fácilmente muchos de los resultados conocidos sobre convergencia de sucesiones de números reales a convergencia de sucesiones de puntos de \mathbb{R}^n . Se tiene, por ejemplo

Teorema 7.2

$$x_k \rightarrow x, y_k \rightarrow y, a_k \xrightarrow{(en \mathbb{R})} a \implies \begin{cases} x_k \pm y_k \rightarrow x \pm y \\ a_k x_k \rightarrow ax \\ \langle x_k, y_k \rangle \rightarrow \langle x, y \rangle \end{cases}$$

(Hay que tener en cuenta que en \mathbb{R}^n , para $n \geq 2$, no hay producto ni cociente de vectores ni tampoco hay definida ninguna relación de orden, por lo que no se pueden dar las nociones de límite superior e inferior.)

Un conjunto S de \mathbb{R}^n se dice que es **acotado** si existe un número real $M > 0$ tal que $|x| < M$ para todo $x \in S$ (o sea, $S \subset B_M(0)$). El **diámetro** de un conjunto A es el número

$$\text{diam}(A) = \sup\{|x - y|; x, y \in A\}$$

Compruébese que A es acotado si y sólo si $\text{diam}(A) < \infty$.

Una sucesión $\{x_k\}$ se dice que es acotada si el conjunto de todos sus elementos es acotado (equivalentemente, si $\{x_i^k\}$ es acotada en \mathbb{R} para cada $i = 1, \dots, n$). Como en \mathbb{R} , se comprueba inmediatamente que una sucesión convergente es acotada.

Una sucesión $\{x_k\}$ se dice que es una sucesión de Cauchy si para todo $\varepsilon > 0$ existe un entero positivo $N = N(\varepsilon)$ tal que, si $p, q \geq N$, se verifica $|x_p - x_q| < \varepsilon$ (lo que es equivalente a decir que las sucesiones de números reales $\{x_i^k\}$, $i = 1, \dots, n$, son, todas ellas, sucesiones de Cauchy). Basándose en el correspondiente resultado en \mathbb{R} (uno de los resultados fundamentales de la recta real) se tiene:

Teorema 7.3 *Una sucesión de puntos de \mathbb{R}^n es convergente si y sólo si es una sucesión de Cauchy.*

La generalización del teorema de Bolzano-Weierstrass tampoco ofrece demasiada dificultad:

Teorema 7.4 (Bolzano-Weierstrass) *Toda sucesión acotada de puntos de \mathbb{R}^n posee una subsucesión convergente.*

Demostración: Sea $\{x_k\}$ la sucesión acotada bajo consideración. Consideramos en primer lugar la sucesión $\{x_1^k\}$; es una sucesión acotada de números reales, por lo que, por el teorema de Bolzano-Weierstrass, posee una subsucesión $\{x_1^{k_m}\}$ tal que $x_1^{k_m} \rightarrow x_1^0$ cuando $m \rightarrow \infty$. Consideramos ahora la sucesión $\{x_2^{k_m}\}$, que también es acotada, con lo que, aplicando de nuevo el teorema de Bolzano-Weierstrass, obtenemos una subsucesión suya —que seguimos llamando igual para no complicar las notaciones— convergente a un x_2^0 . La idea es clara: tenemos ya una subsucesión $\{x_{k_m}\}$ de $\{x_k\}$ que converge en las dos primeras coordenadas; si $n = 2$, la demostración estaría terminada; si no, consideraríamos las terceras coordenadas de los elementos de dicha subsucesión y así, por un proceso de inducción, obtendríamos, después de n extracciones, un subsucesión $\{x_{k_m}\}$ de $\{x_k\}$ convergente a un punto $x_0 = \{x_1^0, \dots, x_n^0\}$ de \mathbb{R}^n .

Se pueden caracterizar mediante sucesiones algunos de los conceptos topológicos introducidos en el apartado anterior:

Proposición 7.4 $x_0 \in \mathbb{R}^n$ es un punto de acumulación del conjunto $A \subset \mathbb{R}^n$ si y sólo si existe una sucesión $\{x_k\}$ de puntos de A tal que $x_k \neq x_0$ para todo $k \in \mathbb{N}$ y $x_k \rightarrow x_0$.

Demostración: (i) Sea x_0 un punto de acumulación de A ; entonces, para todo $k \in \mathbb{N}$ existe $x_k \in B_{1/k}(x_0)$ con $x_k \neq x_0$ y, claramente, $x_k \rightarrow x_0$.

(ii) Recíprocamente, supongamos que $x_k \rightarrow x_0$; entonces, dado $\epsilon > 0$ cualquiera, existe N tal que $x_k \in B_\epsilon(x_0)$ para todo $k \geq N$, es decir, existe $x_k \in B_\epsilon(x_0) \cap A$ con $x_k \neq x_0$, lo que significa que x_0 es punto de acumulación de A .

Como consecuencia, se tienen las siguientes caracterizaciones de la adherencia de un conjunto y de un conjunto cerrado:

Corolario 7.1 (a) $x \in \bar{A}$ si y sólo si existe una sucesión $\{x_k\}$ de puntos de A tal que $x_k \rightarrow x$.

(b) A es cerrado si y sólo si el límite de toda sucesión de puntos de A pertenece a A .

Otra consecuencia interesante del teorema de Bolzano-Weierstrass es la siguiente:

Corolario 7.2 Sean $A_1, A_2, \dots, A_j, \dots$ subconjuntos no vacíos, cerrados y acotados de \mathbb{R}^n tales que $A_1 \supseteq A_2 \supseteq \dots \supseteq A_j \supseteq A_{j+1} \supseteq \dots$. Entonces, el conjunto $\bigcap_{j=1}^{\infty} A_j$ es no vacío.

En efecto, para cada $k \in \mathbb{N}$, elijamos un punto $x_k \in A_k$. La sucesión $\{x_k\}$ está acotada ya que está contenida en el conjunto acotado A_1 ; por tanto, posee una subsucesión $\{x_{k_m}\}$ convergente a un punto x_0 . Puesto que $x_{k_m} \in A_j$ para todo $k_m \geq j$ y A_j es cerrado, se tendrá $x_0 \in A_j$; y como esto es cierto para todo $j \in \mathbb{N}$ se tendrá $x_0 \in \bigcap_{j=1}^{\infty} A_j$.

(EJERCICIO: Pruébese que si, además, $\text{diam}(A_j) \rightarrow 0$ cuando $j \rightarrow \infty$, entonces $\bigcap_{j=1}^{\infty} A_j$ consta de un sólo punto. Esto generaliza el conocido resultado en \mathbb{R} relativo a intervalos encajados, y, de hecho, se puede demostrar directamente sin apelar al teorema de Bolzano-Weierstrass.)

El teorema 7.4 se puede reformular así:

Teorema 7.5 (Bolzano-Weierstrass) Todo conjunto infinito y acotado de \mathbb{R}^n posee al menos un punto de acumulación.

Series en \mathbb{R}^n

En cuanto se tiene la noción de convergencia en \mathbb{R}^n se puede, como en \mathbb{R} , dar el concepto de serie: Dada una sucesión $\{x_k\}$ de puntos de \mathbb{R}^n se forma la sucesión de sumas parciales $\{S_N\}$, donde

$$S_N = x_1 + \dots + x_N$$

Si ocurre que existe

$$\lim_{N \rightarrow \infty} S_N = S \in \mathbb{R}^n$$

se escribe

$$\sum_{k=1}^{\infty} x_k = S$$

o simplemente $\sum x_k = S$, y se dice que la serie $\sum x_k$ converge a S o que la suma de la serie es S .

Es claro que la serie $\sum x_k$ en \mathbb{R}^n es convergente si y sólo si son convergentes todas y cada una de las series numéricas $\sum x_i^k$, $i = 1, \dots, n$.

Se dice que la serie $\sum x_k$ es absolutamente convergente si la serie de números reales $\sum |x_k|$ es convergente. Como en \mathbb{R} , se comprueba que si una serie es absolutamente convergente, entonces es convergente.

Un criterio útil para la convergencia absoluta es, también como en \mathbb{R} , el criterio de comparación de Weierstrass: Si $\sum x_k$ es tal que $|x_k| \leq a_k \in \mathbb{R}$ y la serie de números reales no negativos $\sum a_k$ es convergente, entonces $\sum x_k$ es absolutamente convergente (y, por tanto, convergente). En efecto, se tendrá

$$0 \leq \sum_{k=p+1}^q |x_k| \leq \sum_{k=p+1}^q a_k$$

con lo que, por el criterio de Cauchy aplicado a las sucesiones de sumas parciales $\sum |x_k|$ y $\sum a_k$, resultará que $\sum |x_k|$ es convergente.

7.3 Conjuntos compactos

La propiedad expresada en el teorema de Bolzano-Weierstrass es esencial en la demostración de resultados fundamentales de máximos y mínimos, continuidad uniforme, etc.. Por ello conviene distinguir a los conjuntos de \mathbb{R}^n en los que se puede aplicar dicha propiedad:

Definición 7.8 Un conjunto K de \mathbb{R}^n se dice que es compacto si de toda sucesión $\{x_k\}$ de elementos de K se puede extraer una subsucesión que converge a un punto de K .

Teorema 7.6 Un conjunto K de \mathbb{R}^n es compacto si y sólo si es cerrado y acotado.

Demostración: (i) Supongamos que K es compacto. Sea $\{x_k\}$ una sucesión de puntos de K tal que $x_k \rightarrow x \in \mathbb{R}^n$; por ser K compacto, $\{x_k\}$ posee una subsucesión convergente a un punto de K ; pero toda subsucesión de una sucesión convergente converge al mismo límite que la sucesión completa, por lo que $x \in K$ y K es cerrado por el corolario 7.1. Si no fuese acotado, entonces, para cada $k \in \mathbb{N}$ existiría un punto $x_k \in K$ tal que $|x_k| > k$ y es claro que de esta sucesión $\{x_k\}$ no se podría extraer una subsucesión convergente, lo que contradice la compacidad de K .

(ii) Supongamos ahora que K es cerrado y acotado y sea $\{x_k\}$ una sucesión de elementos de K ; $\{x_k\}$ es acotada, con lo que, por el teorema de Bolzano-Weierstrass, tiene una subsucesión convergente; su límite estará en K por ser K cerrado. En consecuencia, K es compacto.

Hay otra caracterización de los conjuntos compactos que resulta más apropiada para la extensión del concepto a espacios más generales que \mathbb{R}^n . Se basa en la noción de recubrimiento abierto de un conjunto. Una colección \mathfrak{F} de conjuntos se dice que **recubre** al conjunto A o que es un **recubrimiento** de A si $A \subseteq \bigcup_{V \in \mathfrak{F}} V$. Se dice que es un **recubrimiento abierto** si todos los conjuntos $V \in \mathfrak{F}$ son abiertos.

Teorema 7.7 (Heine-Borel) *Un conjunto K de \mathbb{R}^n es compacto si y sólo si de todo recubrimiento abierto de K se puede extraer una subcolección finita que también recubre K .*

Demostración: (i) Es condición necesaria: sea K compacto y sea \mathfrak{F} un recubrimiento abierto de K . Como K es acotado, está contenido en un cubo n -dimensional $C = \{x \in \mathbb{R}^n; -a/2 \leq x_i \leq a/2, i = 1, \dots, n\}$ de lado a y centro 0 (un cuadrado en \mathbb{R}^2 ; véase la figura 7.1).

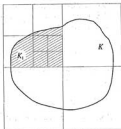


Figura 7.1

C es el producto cartesiano $I_1 \times I_2 \times \dots \times I_n$ de los intervalos $I_i = [-a/2, a/2]$, $i = 1, \dots, n$, o sea, el conjunto de puntos (x_1, \dots, x_n) tales que $x_i \in I_i$ (compruébese que el diámetro de C es $\sqrt{n}a$). Mediante la división de cada I_i en dos intervalos cerrados $I_{i,1} = [-a/2, 0]$ e $I_{i,2} = [0, a/2]$ se obtienen 2^n subcubos $I_{1,\alpha(1)} \times I_{2,\alpha(2)} \times \dots \times I_{n,\alpha(n)}$, donde $\alpha(i)$ vale 1 ó 2, de lado $a/2$ y cuya unión es C .

La demostración procede por reducción al absurdo. Si no se puede extraer de \mathfrak{F} una subcolección finita que recubre a K , entonces, la intersección de K con al menos uno, llamémosle C_1 , de esos 2^n subcubos tampoco puede recubrirse con un número finito de conjuntos de \mathfrak{F} . Apliquemos el mismo argumento a $K_1 = K \cap C_1$, que también es compacto por ser cerrado y acotado, dividiendo a su vez C_1 en 2^n subcubos de lado $a/2^2$. Obtendremos así una sucesión de conjuntos compactos

$$K = K_0 \supset K_1 \supset K_2 \supset \dots \supset K_m \supset \dots$$

ninguno de los cuales puede recubrirse con una subcolección finita de \mathfrak{F} y que satisfacen además $\text{diam}(K_m) \leq (a/2^m)\sqrt{n}$. Por el corolario 7.2 del apartado anterior, existe $x_0 \in \bigcap_{m=0}^{\infty} K_m$. Como $x_0 \in K$, hay al menos un conjunto V de la colección \mathfrak{F} tal que $x_0 \in V$ y, por ser V abierto, existe $\varepsilon > 0$ tal que $B_\varepsilon(x_0) \subset V$. Dado que $\text{diam}(K_m) \rightarrow 0$ cuando $m \rightarrow \infty$, se tendrá que $K_m \subset B_\varepsilon(x_0) \subset V$ para m suficientemente grande, es decir, que tales K_m quedan cubiertos por un único conjunto V de la colección \mathfrak{F} , lo que contradice la hipótesis de que K no se puede recubrir con una subcolección finita de \mathfrak{F} .

(ii) Es condición suficiente: Para ver que K es cerrado y acotado procedemos también por reducción al absurdo. Si K no fuese acotado, del recubrimiento de K formado por las bolas abiertas $B_m(0)$, $m \in \mathbb{N}$, no se podría extraer ninguna subcolección finita que también recubriese a K . Si K no fuese cerrado, entonces existiría un punto $x_0 \in \partial K \setminus K$ y los conjuntos $A_m = \{x \in \mathbb{R}^n; |x - x_0| > 1/m\}$ formarían un recubrimiento abierto de K que tampoco satisfaría la hipótesis.

Observaciones:

1. Analícese la influencia de cada una de las hipótesis mediante los siguientes ejemplos:

- (a) $K = B_1(0)$, \mathfrak{F} formado por $V_m = B_{1-1/m}(0)$, $m = 1, 2, \dots$
- (b) $K = \mathbb{R}^n$, \mathfrak{F} formado por $V_m = B_m(0)$, $m = 1, 2, \dots$
- (c) $K = \bar{B}_1(0)$, \mathfrak{F} formado por $V_0 = \{x \in \mathbb{R}^n; 1 \leq |x| < 2\}$ y $V_m = B_{1-1/m}(0)$, $m = 1, 2, \dots$

2. Tenemos, pues, tres definiciones equivalentes para los conjuntos compactos de \mathbb{R}^n : por la existencia de subsucesiones convergentes, mediante la noción de recubrimientos abiertos o, simplemente, como conjuntos cerrados y acotados.
3. Una aplicación simple y relativamente frecuente del teorema de Heine-Borel es la siguiente: dado un conjunto K y un número $\varepsilon > 0$ es claro que $K \subset \bigcup_{x \in K} B_\varepsilon(x)$; si K es compacto, existirá una cantidad finita de puntos x_1, \dots, x_N de K tales que $K \subset B_\varepsilon(x_1) \cup \dots \cup B_\varepsilon(x_N)$.

7.4 Límite y continuidad de una función

Como en el capítulo 2 ponemos:

Definición 7.9 Sea $f : D \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^m$ y sea x_0 un punto de acumulación de D . Se dice que $f(x)$ tiene un límite $l \in \mathbb{R}^m$ cuando x tiende a x_0 , lo que se expresa por

$$\lim_{x \rightarrow x_0} f(x) = l \quad (7.12)$$

si para todo $\varepsilon > 0$ existe $\delta > 0$ tal que las relaciones

$$x \in D, \quad 0 < |x - x_0| < \delta \quad (7.13)$$

implican $|f(x) - l| < \varepsilon$.

El límite, caso de existir, es único (pruébese como ejercicio). En lugar de (7.12) se escribe también " $f(x) \rightarrow l$ cuando $x \rightarrow x_0$ ".

Como para las funciones de una variable, se puede dar una definición equivalente de límite utilizando sucesiones:

Teorema 7.8 Sea $f : D \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^m$ y sea x_0 un punto de acumulación de D . Las dos afirmaciones siguientes son equivalentes

(i) $\lim_{x \rightarrow x_0} f(x) = l$

(ii) Para toda sucesión $\{x_k\} \subset D$ con $x_k \neq x_0$ para todo k y tal que $x_k \rightarrow x_0$, se tiene $f(x_k) \rightarrow l$.

Demostración: (i) \Rightarrow (ii). Supongamos que $\lim_{x \rightarrow x_0} f(x) = l$ y que $\{x_k\}$ es una sucesión de puntos de D distintos de x_0 que converge a x_0 . Entonces, dado $\varepsilon > 0$ existe, por una parte, $\delta > 0$ tal que, si $x \in D$ y $0 < |x - x_0| < \delta$,

$|f(x) - l| < \varepsilon$ y, por otra, un entero N tal que, si $k \geq N$, $|x_k - x_0| < \delta$. Se sigue de ello que, dado $\varepsilon > 0$, existe un entero N tal que $|f(x_k) - l| < \varepsilon$ para todo $k \geq N$, es decir, que $f(x_k) \rightarrow l$.

(ii) \Rightarrow (i). Vamos a probar que si se satisface (ii) y no se satisface (i) se llega a contradicción. En efecto, si no es cierto que $\lim_{x \rightarrow x_0} f(x) = l$, entonces, existe un $\varepsilon > 0$ tal que, cualquiera que sea $\delta > 0$, existe un punto $x_\delta \neq x_0$ con $0 < |x_\delta - x_0| < \delta$ para el que $|f(x_\delta) - l| \geq \varepsilon$. Tomando los números $\delta_k = 1/k$, $k = 1, 2, \dots$, se obtiene una sucesión $\{x_k\}$ con $x_k \neq x_0$ que converge a x_0 tal que $f(x_k)$ no converge a l , lo que contradice la hipótesis (ii).

De las propiedades de la convergencia de sucesiones o directamente de la definición 7.9, se deducen fácilmente las siguientes propiedades de los límites:

Proposición 7.5 Si $\lim_{x \rightarrow x_0} f(x) = a$, $\lim_{x \rightarrow x_0} g(x) = b$, entonces

$$(i) \quad \lim_{x \rightarrow x_0} (f(x) \pm g(x)) = a \pm b$$

$$(ii) \quad \lim_{x \rightarrow x_0} cf(x) = ca, \text{ para todo } c \in \mathbb{R}$$

$$(iii) \quad \lim_{x \rightarrow x_0} (f(x), g(x)) = (a, b)$$

Asimismo, de los teoremas 7.1 y 7.8 se deduce que, en realidad, las cuestiones de límites de funciones vectoriales se pueden reducir al caso $m = 1$ estudiado en el capítulo 2:

Proposición 7.6 $\lim_{x \rightarrow x_0} f(x) = l = (l_1, \dots, l_m)$ si y sólo si $\lim_{x \rightarrow x_0} f_i(x) = l_i$.

Como para las funciones de \mathbb{R}^2 en \mathbb{R} se tiene:

Proposición 7.7 Si existe $\varepsilon > 0$ tal que todos los puntos de $B_\varepsilon(x_0)$ pertenecen a D excepto quizás el propio x_0 , y

$$l = \lim_{x \rightarrow x_0} f(x)$$

entonces, para todo vector unitario $u \in \mathbb{R}^n$, $|u| = 1$, se tiene

$$\lim_{t \rightarrow 0} f(x_0 + tu) = l \quad (7.14)$$

Los puntos $x_0 + tu$, $t \in \mathbb{R}$, describen la recta de \mathbb{R}^n que pasa por x_0 y tiene a u por vector dirección; el límite (7.14) es el límite de la función de una variable

$$\mathbb{R} \ni t \longmapsto f(x_0 + tu) \in \mathbb{R}^m$$

y lo podemos considerar como el *límite direccional* de f en x_0 según la dirección rectilínea dada por el vector u .

Como vimos en el capítulo 2, el recíproco de esta proposición no es cierto: puede ocurrir que existan y coincidan los límites según cualquier recta que pasa por x_0 y que, sin embargo, no exista $\lim_{x \rightarrow x_0} f(x)$.

Observación: La existencia de límite depende del conjunto D desde el que nos acercamos a x_0 (nótese cómo en la definición de límite aparece claramente la cláusula $x \in D$, lo mismo que en la versión por sucesiones del teorema anterior). Sea, por ejemplo, la función

$$f(x, y) = \frac{xy}{x^2 + y^2} \quad \text{si } (x, y) \neq (0, 0), \quad f(0, 0) = 0$$

y $(x_0, y_0) = (0, 0)$. Si $D = \mathbb{R}^2$, entonces no existe $\lim_{(x, y) \rightarrow (0, 0)} f(x, y)$, como se comprueba fácilmente calculando los límites según las rectas que pasan por el origen. Sin embargo, si $D' = \{(x, y) \in \mathbb{R}^2; x^2 \leq y \leq 2x^2\}$, entonces, para $(0, 0) \neq (x, y) \in D'$ se tiene

$$|f(x, y) - 0| \leq \frac{|x||y|}{x^2 + y^2} \leq \frac{4|x|x^2}{x^2 + x^4} = \frac{4|x|}{1 + x^2} \leq 4|x|$$

de donde se deduce que $\lim_{(x, y) \rightarrow (0, 0)} f(x, y) = 0$. Algo semejante ya se daba en funciones de una variable: puede ocurrir que no exista $\lim_{x \rightarrow x_0} f(x)$ y que sin embargo existan los *límites laterales* $\lim_{x \rightarrow x_0^-} f(x)$ y $\lim_{x \rightarrow x_0^+} f(x)$, que son los límites cuando nos acercamos a x_0 desde, respectivamente, $(-\infty, x_0)$ y (x_0, ∞) , si bien en este caso la terminología es suficientemente explícita y no cabe ninguna ambigüedad en la interpretación de estos límites.

La definición de continuidad también es, naturalmente, la misma que la del capítulo 2:

Definición 7.10 $f : D \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^m$ se dice que es continua en $x_0 \in D$ si

$$\lim_{x \rightarrow x_0} f(x) = f(x_0) \quad (7.15)$$

Como comentábamos allí, en realidad, la condición (7.15) no tiene sentido si x_0 no es punto de acumulación de D , o sea, si se trata de un *punto aislado* de D . Salvamos esta situación excepcional acordando que (7.15) se verifica automáticamente en un punto aislado x_0 , o sea, que $f(x)$ es por definición continua en todo punto aislado de D (lo que, realmente, no choca con la definición 7.9 si tomamos $l = f(x_0)$ y, con un $\delta > 0$ suficientemente

pequeño, permitimos que la condición de aquella sea, para un punto aislado x_0 , " $x \in D$, $0 \leq |x - x_0| < \delta$ implica $|f(x) - l| < \varepsilon$ ": es el propio x_0 el único punto x que interviene en la condición, donde se ha cambiado el signo $<$ por \leq , y ésta resulta trivialmente satisfecha). Con esta aclaración, la definición de continuidad se puede expresar en la forma equivalente (y que, por tanto, podríamos tomar como definición original de continuidad) :

Definición 7.11 $f : D \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^m$ se dice que es continua en $x_0 \in D$ si para todo $\varepsilon > 0$ existe $\delta > 0$ tal que las relaciones

$$x \in D, |x - x_0| < \delta$$

implican

$$|f(x) - f(x_0)| < \varepsilon$$

Se dice que f es continua en un conjunto $A \subseteq D$ si es continua en cada uno de los puntos de A . (En el caso muy especial en que A esté formado únicamente por puntos aislados, cualquier función definida en A es continua en A .)

La continuidad también se puede caracterizar por sucesiones, como sabemos de la teoría de funciones de una variable:

Teorema 7.9 Sea $f : D \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^m$. Las dos afirmaciones siguientes son equivalentes

- (i) f es continua en el punto $x_0 \in D$.
- (ii) Para toda sucesión $\{x_k\} \subset D$ tal que $x_k \rightarrow x_0$, se tiene $f(x_k) \rightarrow f(x_0)$.

La demostración es la misma que la del teorema 7.8. (Si x_0 es un punto aislado de D , las únicas sucesiones que entran en juego son las que, a partir de un cierto índice, se estabilizan en el valor constante x_0 .)

De la proposición 7.5 se deriva:

Proposición 7.8 Si f y g son continuas en x_0 , también lo son $f \pm g$ y cf para todo $c \in \mathbb{R}$

Como en la proposición 7.6, y como consecuencia de ella, la continuidad de una función $f = (f_1, \dots, f_m)$ se puede verificar por la continuidad de sus componentes:

Proposición 7.9 $f = (f_1, \dots, f_m)$ es continua en x_0 si y sólo si cada f_i es continua en x_0 .

En el capítulo 2 enunciamos varios resultados parciales relativos a la **composición** de funciones continuas. Demostraremos a continuación el resultado general del que aquéllos son casos particulares. El planteamiento es el siguiente: si $f: D \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^m$, $g: \Omega \subseteq \mathbb{R}^m \rightarrow \mathbb{R}^p$ y $f(D) \subseteq \Omega$, entonces está definida la **aplicación compuesta** $g \circ f: D \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^p$ por $(g \circ f)(x) = g(f(x))$. Pues bien, se tiene:

Teorema 7.10 *Supongamos que $f: D \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^m$, $g: \Omega \subseteq \mathbb{R}^m \rightarrow \mathbb{R}^p$ y $f(D) \subseteq \Omega$. Si f es continua en $x_0 \in D$ y g es continua en $y_0 = f(x_0)$, entonces $g \circ f$ es continua en x_0 .*

Demostración: Dado $\varepsilon > 0$, hemos de encontrar $\delta > 0$ de modo que $|(g \circ f)(x) - (g \circ f)(x_0)| < \varepsilon$ siempre que $x \in D$ verifique $|x - x_0| < \delta$. Por la continuidad de g en $y_0 = f(x_0)$, existe $\eta > 0$ tal que

$$y \in \Omega, |y - f(x_0)| < \eta \implies |g(y) - g(f(x_0))| < \varepsilon \quad (7.16)$$

A su vez, por la continuidad de f en x_0 , existe $\delta > 0$ tal que $|f(x) - f(x_0)| < \eta$ siempre que $x \in D$ y $|x - x_0| < \delta$. Podemos, entonces, sustituir $y = f(x)$ en (7.16) y concluir que, como queríamos,

$$x \in D, |x - x_0| < \delta \implies |g(f(x)) - g(f(x_0))| < \varepsilon$$

Observación: En correspondencia con la observación precedente acerca de los límites, cabe decir que la noción de continuidad depende del conjunto D en que suponemos definida la función y, así, la función

$$f(x, y) = \frac{xy}{x^2 + y^2} \quad \text{si } (x, y) \neq (0, 0), \quad f(0, 0) = 0$$

no es continua en el punto $(0, 0)$ desde $D = \mathbb{R}^2$ pero sí lo es desde $D' = \{(x, y) \in \mathbb{R}^2; x^2 \leq y \leq 2x^2\}$. Desde luego, lo más frecuente es que, en el estudio de una función f , el conjunto D en el que se supone que está actuando sea el propio dominio de definición de f y podría pensarse que la restricción a un conjunto D' como el anterior es una pequeña "trampa" para fabricar un límite para una función que realmente no lo tiene o hacer continua una función que realmente no lo es. Piénsese, sin embargo, que dicha restricción podría tener perfecto sentido en el siguiente problema: Minimizar f sujeta a $y \geq x^2$, $y \leq 2x^2$. Nótese, por otra parte, que tomando $D = \mathbb{R}^2$ como dominio de f , f no es continua en D' (aunque sí lo es desde D') ya que no es continua en cada punto de D' (desde D). En otras palabras, y para precisar las ideas, si denotamos de manera distinta a ambas funciones,

llamando f a la función dada con dominio \mathbb{R}^2 y g a su restricción a D' , podemos decir lo siguiente

g es continua en D'
 f , que es una extensión de g , no es continua en D'

O sea: Que g sea continua en D' no implica que admita extensiones que sí lo sean.

Funciones continuas y conjuntos compactos

Las funciones continuas tienen la importante propiedad de que transforman conjuntos compactos en conjuntos compactos:

Teorema 7.11 Sea K un conjunto compacto de \mathbb{R}^n y sea $f: K \rightarrow \mathbb{R}^m$ continua en K . Entonces, $f(K)$ es un conjunto compacto de \mathbb{R}^m .

Demostración: Sea $\{y_k\}$ una sucesión cualquiera de puntos de $f(K)$; hay que probar que posee una subsucesión convergente a un punto de $f(K)$. Para cada y_k existe un $x_k \in K$ tal que $y_k = f(x_k)$; como, por hipótesis, K es compacto, $\{x_k\}$ tiene una subsucesión $\{x_{k_m}\}$ convergente a un punto $x_0 \in K$ y, entonces, por la continuidad de f se tendrá, de acuerdo con el teorema 7.9, que $y_{k_m} = f(x_{k_m}) \rightarrow f(x_0) \in f(K)$.

Cuando el teorema anterior se aplica a funciones con valores en \mathbb{R} (o sea, a funciones *escalares*) resulta el **teorema de Weierstrass** sobre existencia de máximos y mínimos:

Teorema 7.12 Sea K un conjunto cerrado y acotado de \mathbb{R}^n y sea $f: K \rightarrow \mathbb{R}$ continua en K . Entonces, f alcanza en K su valor máximo y su valor mínimo, es decir, existen $x^* \in K$ y $x^{**} \in K$ tales que

$$f(x^*) \leq f(x) \leq f(x^{**}) \text{ para todo } x \in K$$

Demostración: Por el teorema anterior, $f(K)$ es un conjunto cerrado y acotado de la recta \mathbb{R} ; por ser acotado, $f(K)$ tiene un supremo M y un ínfimo m , y estos números han de pertenecer a $f(K)$ porque también es cerrado; en consecuencia, existen $x^* \in K$ y $x^{**} \in K$ tales que $f(x^*) = m$, $f(x^{**}) = M$. (Véase el problema 7.)

NOTA. La imagen por una función continua de un conjunto cerrado no tiene por qué ser cerrada ni la de un conjunto acotado tiene por qué ser acotada: considérense los ejemplos en una variable:

$$(i) \quad f(x) = \frac{x^2}{x^2 + 1}, \quad f(\mathbb{R}) = [0, 1)$$

$$(ii) \quad f(x) = \frac{1}{x}, \quad f((0, 1]) = [1, \infty)$$

Continuidad uniforme

Si f es continua en un conjunto A , el número $\delta > 0$ correspondiente, en la definición de continuidad, a un $\varepsilon > 0$ dado, dependerá, en general, del punto $x_0 \in A$ considerado. Cuando para cada $\varepsilon > 0$ existe un $\delta = \delta(\varepsilon) > 0$ con el que se satisface la definición de continuidad para *todo* $x_0 \in A$, se dice que f es **uniformemente continua** en A :

Definición 7.12 Una función $f: A \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^m$ se dice que es **uniformemente continua** en A si para todo $\varepsilon > 0$ existe $\delta > 0$ tal que las relaciones

$$x, y \in D, |x - y| < \delta$$

implican

$$|f(x) - f(y)| < \varepsilon$$

La propiedad de continuidad uniforme es estrictamente más fuerte que la continuidad, es decir, existen funciones continuas que no son uniformemente continuas, como muestran los ejemplos, seguramente ya conocidos por el lector, de las funciones $f(x) = 1/x$ en $(0, 1]$ y $g(x) = x^2$ en \mathbb{R} (compruébense los detalles como ejercicio). Ahora bien, en un conjunto compacto las dos propiedades son equivalentes:

Teorema 7.13 Sea K un conjunto compacto de \mathbb{R}^n y sea $f: K \rightarrow \mathbb{R}^m$ continua en K . Entonces, f es uniformemente continua en K .

Demostración: Supongamos que f es continua pero no uniformemente continua en K . Entonces, existe algún $\varepsilon > 0$ tal que para cualquier $\delta > 0$ queelijamos hay dos puntos $x, y \in K$ que verifican

$$|x - y| < \delta \text{ y } |f(x) - f(y)| \geq \varepsilon$$

Tomemos, en particular, $\delta_k = 1/k$, $k = 1, 2, \dots$; entonces, para cada $k \in \mathbb{N}$ existen x_k, y_k tales que

$$|x_k - y_k| < 1/k \text{ y } |f(x_k) - f(y_k)| \geq \varepsilon \quad (7.17)$$

Por ser K compacto, la sucesión $\{x_k\}$ tiene una subsucesión $\{x_{k_m}\}$ convergente a un punto $x \in K$. De la primera desigualdad de (7.17) se deduce —teniendo en cuenta la desigualdad $|x - y_{k_m}| \leq |x - x_{k_m}| + |x_{k_m} - y_{k_m}|$ — que la subsucesión $\{y_{k_m}\}$ converge también a x , y de la segunda, que $f(x_{k_m})$ y $f(y_{k_m})$ no pueden converger al mismo punto; pero esto contradice la continuidad de f , la cual implica que ambas sucesiones han de converger a $f(x)$.

Llegamos, pues, a una contradicción con la hipótesis y la función f ha de ser uniformemente continua.

Por el interés que tiene en contextos más generales el argumento utilizado, merece la pena hacer también la demostración utilizando el teorema de Heine-Borel con la idea que se apunta en la observación 3) con que finaliza el apartado anterior:

Sea $\varepsilon > 0$. Por la hipótesis de continuidad se tiene que para todo $\xi \in K$ existe $\delta = \delta(\xi) > 0$ de modo que

$$x \in K, |x - \xi| < \delta(\xi) \implies |f(x) - f(\xi)| < \varepsilon/2 \quad (7.18)$$

La familia de bolas $B_{\delta(\xi)/2}(\xi)$, $\xi \in K$, constituye un recubrimiento abierto de K ; por ser éste compacto, existe un número finito ξ_1, \dots, ξ_N de puntos de K tales que

$$K \subset B_{\delta(\xi_1)/2}(\xi_1) \cup \dots \cup B_{\delta(\xi_N)/2}(\xi_N)$$

Sea $\delta = \min\{\delta(\xi_1)/2, \dots, \delta(\xi_N)/2\}$ y sean $x, y \in K$ tales que $|x - y| < \delta$; entonces, existe algún ξ_j , $1 \leq j \leq N$, tal que $x \in B_{\delta(\xi_j)/2}(\xi_j)$, es decir $|x - \xi_j| < \delta(\xi_j)/2 < \delta(\xi_j)$; además

$$|y - \xi_j| \leq |y - x| + |x - \xi_j| < \delta + \delta(\xi_j)/2 < \delta(\xi_j)$$

Por la continuidad de f en ξ_j , se tiene

$$|f(x) - f(\xi_j)| < \varepsilon/2 \text{ y } |f(y) - f(\xi_j)| < \varepsilon/2$$

y, en consecuencia,

$$|f(x) - f(y)| \leq |f(x) - f(\xi_j)| + |f(y) - f(\xi_j)| < \varepsilon/2 + \varepsilon/2 = \varepsilon$$

como queríamos demostrar (parecería, en principio, que se podría trabajar con el recubrimiento formado por las bolas $B_{\delta(\xi)}(\xi)$, $\xi \in K$, pero enseguida se ve que las cosas se complican y hay que hacer la corrección a $B_{\delta(\xi)/2}(\xi)$, $\xi \in K$).

Un tipo de funciones uniformemente continuas que aparecen en diversos contextos en el análisis matemático son las funciones *lipschitzianas*:

Definición 7.13 Una función f se dice que es *lipschitziana* o que satisface una *condición de Lipschitz* en el conjunto $A \subseteq \text{Dom } f$ si existe una constante $k > 0$ tal que

$$|f(x) - f(y)| \leq k|x - y|$$

para todo par de puntos $x, y \in A$. Si, en particular, $k < 1$, se dice que f es *contractiva*.

Una función lipschitziana tiene evidentemente la propiedad de continuidad uniforme pues para cada $\varepsilon > 0$ basta tomar $\delta = \varepsilon/k$. El recíproco no es cierto; por ejemplo, la función de una variable $f(x) = x^{2/3}$ es continua en todo \mathbb{R} y, por tanto, uniformemente continua en cualquier intervalo compacto $[a, b]$; sin embargo, no satisface ninguna condición de Lipschitz en el intervalo $[0, 1]$; en efecto, si así fuese, existiría una constante $k > 0$ tal que

$$|f(x) - f(0)| \leq k|x - 0| = k|x|$$

para todo $x \in [0, 1]$. Pero esto implicaría

$$\frac{1}{|x|^{1/3}} \leq k$$

para todo $x \in (0, 1]$, lo que, claramente, no es cierto: basta tomar x suficientemente pequeño para contradecirlo.

Convergencia uniforme

En el curso de funciones de una variable se estudian sucesiones y series de funciones y su convergencia (véase, por ejemplo, [10]). Las ideas y resultados se extienden sin dificultad a funciones de varias variables con valores vectoriales: Sea $\{f_k\}$ una sucesión de funciones $f_k : D \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^m$. Para cada $x \in D$, $\{f_k(x)\}$ es una sucesión de puntos de \mathbb{R}^m ; si $\{f_k(x)\}$ converge, el límite dependerá, en general, de x y lo denotamos por $f(x)$. Quedará así definida una función $x \mapsto f(x)$ en el conjunto $A \subseteq D$ de puntos para los que $\{f_k(x)\}$ es convergente y la situación se describe diciendo que $\{f_k\}$ converge punto a punto (o puntualmente) a f en A . Con precisión:

Definición 7.14 Sea $\{f_k\}$ una sucesión de funciones $f_k : D \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^m$. Se dice que $\{f_k\}$ converge punto a punto (o puntualmente) a la función f en $A \subseteq D$ si para todo $\varepsilon > 0$ y todo $x \in A$ existe un número natural $N(\varepsilon, x)$ tal que

$$|f_k(x) - f(x)| < \varepsilon \text{ para todo } k \geq N(\varepsilon, x)$$

(Por ejemplo, la sucesión de funciones $f_k : [0, \infty) \rightarrow \mathbb{R}$ definidas por $f_k(x) = x^k$ convergen en $A = [0, 1]$ a la función

$$f(x) = 0 \text{ si } 0 \leq x < 1, f(1) = 1$$

si $x > 1$, $\{x^k\}$ diverge.)

Si en la definición anterior ocurre que para cualquier $\varepsilon > 0$ el número N subordinado vale para todo punto $x \in A$, se dice que las funciones convergen uniformemente:

Definición 7.15 Sea $\{f_k\}$ una sucesión de funciones $f_k : A \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^m$. Se dice que $\{f_k\}$ converge uniformemente en A a la función f si para todo $\varepsilon > 0$ existe un número natural $N(\varepsilon)$ tal que

$$|f_k(x) - f(x)| < \varepsilon \text{ para todo } k \geq N(\varepsilon) \text{ y todo } x \in A$$

Es fácil probar que la convergencia uniforme es equivalente a propiedad de Cauchy uniforme:

Definición 7.16 Sea $\{f_k\}$ una sucesión de funciones $f_k : A \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^m$. Se dice que $\{f_k\}$ tiene la propiedad de Cauchy uniforme en A si para todo $\varepsilon > 0$ existe un número natural $N(\varepsilon)$ tal que

$$|f_p(x) - f_q(x)| < \varepsilon \text{ para todo } p \geq N(\varepsilon), q \geq N(\varepsilon) \text{ y todo } x \in A$$

(Obsérvese que esta propiedad garantiza la existencia de la función límite ya que los puntos $\{f_k(x)\}$ forman para cada $x \in A$ una sucesión de Cauchy, por tanto, convergente, de puntos de \mathbb{R}^m .)

Si hay convergencia uniforme, la continuidad de los términos de la sucesión se hereda por la función límite:

Teorema 7.14 Sea $\{f_k\}$ una sucesión de funciones $f_k : A \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^m$ continuas en A . Si $f_k \rightarrow f$ uniformemente en A , f es continua en A .

Demostración: Sea $x_0 \in A$ cualquiera. Por la desigualdad triangular se tiene para todo $k \in \mathbb{N}$:

$$|f(x) - f(x_0)| \leq |f(x) - f_k(x)| + |f_k(x) - f_k(x_0)| + |f_k(x_0) - f(x_0)| \quad (7.19)$$

Dado $\varepsilon > 0$ podemos elegir, por la convergencia uniforme, un $N \in \mathbb{N}$ suficientemente grande tal que

$$|f(x) - f_N(x)| < \varepsilon/3 \text{ para todo } x \in A \quad (7.20)$$

En particular,

$$|f_N(x_0) - f(x_0)| < \varepsilon/3 \quad (7.21)$$

Por ser f_N continua, existe $\delta > 0$ tal que

$$|f_N(x) - f_N(x_0)| < \varepsilon/3 \text{ para todo } x \in A \cap B_\delta(x_0) \quad (7.22)$$

Utilizando (7.20), (7.21) y (7.22) en (7.19), con $k = N$, se tiene que, para todo $x \in A \cap B_\delta(x_0)$, $|f(x) - f(x_0)| < \varepsilon/3 + \varepsilon/3 + \varepsilon/3 = \varepsilon$.

(La convergencia de las funciones continuas $f_k(x) = x^k$ en $A = [0, 1]$ no es uniforme ya que la función límite no es continua.)

Como en el apartado 2, también se puede dar el concepto de *serie de funciones* a partir de una sucesión de funciones $f_k : A \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^m$ y de la *sucesión de sumas parciales*

$$S_N(x) = f_1(x) + \dots + f_N(x)$$

Para cada $x \in A$, $\sum_{k=1}^{\infty} f_k(x)$ es un serie de puntos de \mathbb{R}^m ; si converge para todo $x \in A$, o sea, si converge a un punto $S(x)$ la sucesión de sumas parciales $\{S_N(x)\}$, se dice que la serie $\sum_{k=1}^{\infty} f_k$ converge punto a punto en A . Si $\{S_N\}$ converge uniformemente en A , se dice que $\sum_{k=1}^{\infty} f_k$ converge uniformemente en A . El límite de la sucesión de sumas parciales es la suma de la serie y se escribe $S = \sum_{k=1}^{\infty} f_k$.

Del teorema anterior se deduce inmediatamente:

Teorema 7.15 Sea $\{f_k\}$ una sucesión de funciones $f_k : A \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^m$ continuas en A . Si $\sum_{k=1}^{\infty} f_k$ converge uniformemente en A , entonces la suma de la serie es una función continua en A .

Combinando este teorema con el criterio de comparación de Weierstrass que mencionábamos en el apartado 7.2, se tiene el siguiente resultado práctico:

Corolario 7.3 Sea $\{f_k\}$ una sucesión de funciones $f_k : A \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^m$ continuas en A . Si existe una sucesión de números reales $\{a_k\}$ tales que $|f_k(x)| \leq a_k$ para todo $x \in A$ y la serie $\sum a_k$ es convergente, entonces la serie de funciones $\sum f_k$ converge uniformemente en A y la suma de la serie es una función continua en A .

Caracterización de Hausdorff de las funciones continuas

Comencemos con el caso más simple de una función definida en todo \mathbb{R}^n . La propiedad de ser $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ continua en $x_0 \in \mathbb{R}^n$ puede expresarse así: para todo $\varepsilon > 0$ existe un $\delta > 0$ tal que

$$B_\delta(x_0) \subset f^{-1}(B_\varepsilon(f(x_0))) \quad (7.23)$$

(recuérdese que para cualquier aplicación $f : A \rightarrow B$, la imagen inversa de un subconjunto V de B es el subconjunto de A :

$$f^{-1}(V) = \{x \in A; f(x) \in V\})$$

Teniendo en cuenta la definición de **entorno** de un punto vemos que la condición (7.23) es equivalente a:

$$\boxed{\begin{array}{l} f: \mathbb{R}^n \longrightarrow \mathbb{R}^m \text{ es continua en } x_0 \in \mathbb{R}^n \text{ si} \\ \text{para todo entorno } V \text{ de } f(x_0), f^{-1}(V) \text{ es un entorno de } x_0 \end{array}} \quad (7.24)$$

La continuidad de $f: \mathbb{R}^n \longrightarrow \mathbb{R}^m$ en todo \mathbb{R}^n puede también caracterizarse en términos de conjuntos abiertos y cerrados:

Teorema 7.16 *Dada $f: \mathbb{R}^n \longrightarrow \mathbb{R}^m$, las tres afirmaciones siguientes son equivalentes:*

- (i) f es continua en \mathbb{R}^n ,
- (ii) Para todo conjunto abierto $V \subset \mathbb{R}^m$, $f^{-1}(V)$ es un conjunto abierto de \mathbb{R}^n ,
- (iii) Para todo conjunto cerrado $G \subset \mathbb{R}^m$, $f^{-1}(G)$ es un conjunto cerrado de \mathbb{R}^n .

Demostración: (i) \Rightarrow (ii) Sea $V \subset \mathbb{R}^m$ abierto y sea $x_0 \in f^{-1}(V)$ cualquiera; entonces, $f(x_0) \in V$ y V es entorno de $f(x_0)$, ya que un conjunto abierto si y sólo si es entorno de todos sus puntos; por (7.24), $f^{-1}(V)$ es entorno de x_0 , y, al ser $x_0 \in A$ arbitrario, eso quiere decir que $f^{-1}(V)$ es abierto.

(ii) \Rightarrow (i) Sea $x_0 \in \mathbb{R}^n$ cualquiera; hay que probar que f es continua en x_0 ; sea $\varepsilon > 0$ dado; el conjunto $B_\varepsilon(f(x_0))$ es abierto, luego, por (ii), $f^{-1}(B_\varepsilon(f(x_0)))$ es un abierto que contiene a x_0 ; por tanto, existe $\delta > 0$ tal que $B_\delta(x_0) \subset f^{-1}(B_\varepsilon(f(x_0)))$, pero esto significa que f es continua en x_0 , como expresa (7.23).

(ii) \Leftrightarrow (iii) Esto se comprueba fácilmente tomando complementarios.

En el caso general de una función $f: D \subset \mathbb{R}^n \longrightarrow \mathbb{R}^m$, hay que comenzar por definir los conceptos de entorno, conjunto abierto y conjunto cerrado relativos a un subconjunto D de \mathbb{R}^n :

- Dado un punto $x_0 \in D$, un *entorno de x_0 relativo a D* es un conjunto $U = D \cap W$, donde W es un entorno de x_0 (en \mathbb{R}^n).
- Un conjunto $A \subseteq D$ se dice que es *abierto relativo a D* si existe un conjunto abierto $U \subseteq \mathbb{R}^n$ tal que $A = D \cap U$. La familia de todos los conjuntos abiertos relativos a D constituye la *topología relativa* de D .
- Un conjunto $A \subseteq D$ se dice que es *cerrado relativo a D* si existe un conjunto cerrado $F \subseteq \mathbb{R}^n$ tal que $A = D \cap F$.

Compruébense los siguientes hechos:

- (a) $U \subseteq D$ es entorno de x_0 relativo a D si existe $\delta > 0$ tal que $B_\delta(x_0) \cap D \subseteq U$.
- (b) $A \subseteq D$ es abierto relativo a D si y sólo si para todo $x \in A$ existe $\delta(x) > 0$ tal que $B_{\delta(x)}(x) \cap D \subseteq A$ (o sea, si es entorno relativo de todos sus puntos) (Indicación: Si $A = D \cap U$, entonces cualquier $x \in A$ pertenece a U , abierto de \mathbb{R}^n , con lo que existe $\delta(x) > 0$ tal que $B_{\delta(x)}(x) \subset U$ y $B_{\delta(x)}(x) \cap D \subseteq A$. Para el recíproco, considérese el abierto $U = \bigcup \{B_{\delta(x)}(x); x \in A\}$).
- (c) $A \subseteq D$ es cerrado relativo a D si $D \setminus A$ es abierto relativo a D .

Con estos conceptos, la afirmación (7.24) y el teorema 7.16 valen para funciones $f: D \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$ cambiando los términos entorno de x_0 , abierto de \mathbb{R}^n (\mathbb{R}^m) y cerrado de \mathbb{R}^n (\mathbb{R}^m) por los correspondientes conceptos relativos a D ($f(D)$). (Complétense los detalles como ejercicio).

Obsérvese que la definición de continuidad en un punto (lo mismo que la de límite), que es relativa al conjunto D en el que suponemos definida la función, queda perfectamente precisada mediante estos conceptos topológicos relativos a D (reinterpretese, en particular, el convenio anteriormente indicado de considerar que cualquier función $f: D \rightarrow \mathbb{R}^n$ es continua en un punto aislado de D , así como las distinciones que establecimos en su momento a propósito de la continuidad desde un conjunto y la continuidad en un conjunto. También son ejercicios interesantes releer a la luz de los resultados anteriores la demostración de que la composición de funciones continuas es continua y rehacer la demostración de que la imagen de un compacto por una función continua es un compacto utilizando recubrimientos).

Corolario 7.4 Sea $f: D \subset \mathbb{R}^n \rightarrow \mathbb{R}$ continua y sea $c \in \mathbb{R}$. Los conjuntos $\{x \in D; f(x) < c\}$ y $\{x \in D; f(x) > c\}$ son abiertos y los conjuntos $\{x \in D; f(x) \leq c\}$, $\{x \in D; f(x) \geq c\}$ y $\{x \in D; f(x) = c\}$ son cerrados (relativos a D).

Este corolario (véase también el problema 2.10) permite establecer fácilmente si son abiertos o cerrados conjuntos definidos en \mathbb{R}^n mediante igualdades o desigualdades. Por ejemplo, el conjunto de \mathbb{R}^2

$$A = \{(x, y); xy \geq 1, y \leq 4 - x^2\}$$

es cerrado por ser intersección de los conjuntos

$$A_1 = \{(x, y); f(x, y) \geq 1\}, \quad A_2 = \{(x, y); g(x, y) \leq 4\}$$

ambos cerrados por ser continuas en todo \mathbb{R}^2 las funciones $f(x, y) = xy$, $g(x, y) = x^2 + y$.

La caracterización (7.24) será la que se utilice como *definición* de continuidad en el contexto general de los espacios topológicos y de ella derivará como aquí la afirmación del teorema 7.16 (véase la nota 1 de final de capítulo).

7.5 Normas en \mathbb{R}^n

Cualquier aplicación

$$N : \mathbb{R}^n \longrightarrow [0, \infty)$$

que satisfaga, como la norma euclídea $x \mapsto |x| = (x_1^2 + \dots + x_n^2)^{\frac{1}{2}}$, las propiedades

$$N(x) \geq 0 \text{ para todo } x \in \mathbb{R}^n \text{ y } N(x) = 0 \text{ si y sólo si } x = 0 \quad (7.25)$$

$$N(x + y) \leq N(x) + N(y) \text{ (desigualdad triangular)} \quad (7.26)$$

$$N(cx) = |c| N(x), \quad c \in \mathbb{R} \quad (7.27)$$

se dice que es una **norma** en \mathbb{R}^n . Por ejemplo, se comprueba fácilmente que también son normas en \mathbb{R}^n :

$$|x|_1 = |x_1| + \dots + |x_n| \quad (7.28)$$

y

$$|x|_\infty = \max\{|x_1|, \dots, |x_n|\} \quad (7.29)$$

A una norma $N : \mathbb{R}^n \longrightarrow [0, \infty)$ está asociada una **distancia** entre puntos de \mathbb{R}^n dada por

$$d_N(x, y) = N(x - y) \quad (7.30)$$

que satisface también las propiedades (7.6), (7.7) y (7.8) de la distancia euclídea. Con ella se pueden definir las correspondientes bolas abiertas

$$\{x \in \mathbb{R}^n ; N(x - x_0) < \varepsilon\} \quad (7.31)$$

de centro x_0 y radio ε (en la figura 7.2 se muestra la bola unidad correspondiente a la norma euclídea, $|\cdot|_1$ y $|\cdot|_\infty$).

Y con las bolas abiertas, se definen los conceptos de entorno, conjunto abierto (y la familia de conjuntos abiertos será la *topología* definida por la

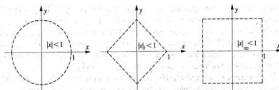


Figura 7.2

norma N), interior, frontera, conjunto cerrado, convergencia de sucesiones, conjunto compacto, función continua,.... Pues bien, resulta que todas estas nociones *topológicas* son **independientes** de la norma en \mathbb{R}^n que se utilice para definir los ε -entornos (7.31) y eso permitirá, a la hora de trabajar con dichas nociones, utilizar en cada caso la norma que resulte más cómoda. La razón de ello estriba en el resultado fundamental, que enseguida demostraremos, de que en \mathbb{R}^n todas las normas son equivalentes.

Definición 7.17 Dos normas N_1 y N_2 en \mathbb{R}^n son **equivalentes** si existen dos constantes $A > 0$, $B > 0$ tales que

$$A N_2(x) \leq N_1(x) \leq B N_2(x) \quad (7.32)$$

para todo $x \in \mathbb{R}^n$

((7.32) equivale a $B^{-1} N_1(x) \leq N_2(x) \leq A^{-1} N_1(x)$, con lo que se trata efectivamente de una relación de equivalencia ya que también es claro que si N_1 es equivalente a N_2 y N_2 lo es a N_3 , entonces N_1 es equivalente a N_3).

Se tiene, por ejemplo,

$$|x|_1 \leq \sqrt{n} |x| \leq \sqrt{n} |x|_1 \quad (7.33)$$

(véase el apartado 1.4) y también

$$|x|_\infty \leq |x| \leq |x|_1 \leq n |x|_\infty \quad (7.34)$$

Si dos normas N_1 y N_2 son equivalentes, entonces, dado el conjunto $\{x \in \mathbb{R}^n; N_1(x - x_0) < \varepsilon\}$, se tiene, por (7.32), que

$$\{x \in \mathbb{R}^n; N_2(x - x_0) < \varepsilon/B\} \subset \{x \in \mathbb{R}^n; N_1(x - x_0) < \varepsilon\}$$

es decir, que todo ε -entorno para la norma N_1 contiene un ε -entorno para la norma N_2 ; en consecuencia, si A es un conjunto abierto en la topología

definida por N_1 , también lo será en la topología definida por N_2 , y, como es evidente, también es cierto el recíproco. Y así ocurre para las demás nociones antes mencionadas ($\{x_k\}$ converge según la distancia asociada a N_1 si y sólo si converge según la distancia asociada a N_2 , etc.), las cuales, en último término, se remiten, en su definición o caracterización, a los conjuntos abiertos, y éstos son los mismos cualquiera que sea la norma considerada.

Para demostrar la equivalencia de todas las normas definidas sobre \mathbb{R}^n , basta demostrar la equivalencia de una cualquiera con la norma euclídea:

Proposición 7.10 Sea $N : \mathbb{R}^n \rightarrow [0, \infty)$ una norma cualquiera. Entonces existen dos constantes $A > 0$, $B > 0$ tales que

$$A |x| \leq N(x) \leq B |x| \quad (7.35)$$

Demostración: Se tiene

$$\begin{aligned} N(x) &= N\left(\sum_{i=1}^n x_i e_i\right) \leq (\text{por (7.26)}) \leq \sum_{i=1}^n |x_i| N(e_i) \leq \\ &\leq M |x|_1 \leq (\text{por (7.33)}) \leq M \sqrt{n} |x| \end{aligned}$$

siendo $\{e_1, \dots, e_n\}$ la base canónica de \mathbb{R}^n y $M = \max\{N(e_i)\}$. Por otra parte, por la desigualdad triangular

$$|N(x) - N(y)| \leq N(x - y) \leq M \sqrt{n} |x - y|$$

de donde se deduce que $N : \mathbb{R}^n \rightarrow \mathbb{R}$ es (uniformemente) continua. En consecuencia, alcanzará un valor máximo B y un valor mínimo A en el conjunto compacto $S_1 = \{x \in \mathbb{R}^n; |x| = 1\}$. Si $x \in \mathbb{R}^n$, $x \neq 0$, se tendrá $x/|x| \in S_1$, con lo que

$$A \leq N(x/|x|) = \frac{1}{|x|} N(x) \leq B$$

y en consecuencia

$$A |x| \leq N(x) \leq B |x|$$

desigualdad que obviamente se satisface también para $x = 0$.

7.6 Aplicaciones lineales

Las aplicaciones lineales de \mathbb{R}^n en \mathbb{R}^m que se estudian en Álgebra lineal son, de hecho, las funciones vectoriales más simples. Como se sabe, a una aplicación

lineal $L: \mathbb{R}^n \rightarrow \mathbb{R}^m$ está asociada una única matriz $m \times n$

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \quad (7.36)$$

tal que, si $y = Lx$,

$$y_i = \sum_{j=1}^n a_{ij} x_j, \quad i = 1, \dots, m \quad (7.37)$$

En particular, $L(e_j) = (a_{1j}, \dots, a_{mj})$ (la imagen por L del vector j -ésimo de la base canónica es la columna j -ésima de la matriz; (7.37) expresa simplemente el producto usual de una matriz por un vector columna). Se puede, pues, identificar el conjunto $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ de aplicaciones lineales de \mathbb{R}^n en \mathbb{R}^m con el conjunto $\mathcal{M}_{m \times n}$ de matrices $m \times n$, el cual, a su vez, se identifica de manera natural al espacio $\mathbb{R}^{m \times n}$ (contemplando la matriz $A = (a_{ij})$ como una $(m \times n)$ -upla ordenada de números reales, colocando, por ejemplo, la segunda columna después de la primera y así sucesivamente).

Las aplicaciones lineales son uniformemente continuas:

Proposición 7.11 Sea $L \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$; para todo $x \in \mathbb{R}^n$ se tiene

$$|Lx| \leq M|x| \quad \text{con} \quad M = \left(\sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^2 \right)^{1/2} \quad (7.38)$$

En efecto, basta aplicar la desigualdad de Cauchy-Schwartz a la suma de (7.37):

$$|y_i|^2 \leq \left(\sum_{j=1}^n |a_{ij}|^2 \right) \left(\sum_{j=1}^n |x_j|^2 \right)$$

y sumar en i de 1 a n .

De (7.38) se deduce, por la linealidad de L ,

$$|Lx - Ly| = |L(x - y)| \leq M|x - y| \quad (7.39)$$

y de aquí, que L es lipschitziana y, por tanto, uniformemente continua en \mathbb{R}^n .

La identificación de $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ y $\mathcal{M}_{m \times n}$ con $\mathbb{R}^{m \times n}$ nos permite hablar de la *norma de una aplicación lineal* $L \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ (o *norma de una matriz*

$m \times n$), y todas las normas de L son equivalentes (por ejemplo, el número M de (7.38) es la norma euclídea de L , o de su matriz asociada $A = (a_{ij})$, en $\mathbb{R}^{m \times n}$).

Fijadas unas determinadas normas en \mathbb{R}^n y \mathbb{R}^m , tiene especial interés la **norma uniforme** de una aplicación lineal $L: \mathbb{R}^n \rightarrow \mathbb{R}^m$, definida por la magnitud

$$\|L\| = \max\{|Lx|; |x| \leq 1\} \quad (7.40)$$

Está bien definida pues $\{x \in \mathbb{R}^n; |x| \leq 1\}$ es compacto y L , como acabamos de ver, es continua (con lo que se puede aplicar el teorema de Weierstrass a la aplicación compuesta $x \mapsto Lx \mapsto |Lx| \in \mathbb{R}$). Compruébese como ejercicio que, efectivamente, satisface las propiedades (7.25), (7.26) y (7.27) de norma en $\mathbb{R}^{m \times n}$. Se tiene además

$$|Lx| \leq \|L\| |x| \quad \text{para todo } x \in \mathbb{R}^n \quad (7.41)$$

(si $x = 0$, entonces $|Lx| = 0 = \|L\| |x|$ y si $x \neq 0$, entonces $|x| \neq 0$ y $|x|^{-1}x$ tiene norma 1, con lo que, por la definición (7.40), resulta $\|L\| \geq |L(|x|^{-1}x)| = |x|^{-1}|Lx|$).

$\|L\|$ es la menor constante k para la que se verifica

$$|Lx| \leq k |x| \quad \text{para todo } x \in \mathbb{R}^n$$

(el valor M de (7.38) no es, en general, el óptimo; por ejemplo, si L es la matriz identidad en \mathbb{R}^2 , se tiene $M = \sqrt{2} > 1 = \|L\|$).

También se tiene

$$\|L\| = \max\{|Lx|; |x| = 1\} = \sup \left\{ \frac{|Lx|}{|x|}; x \in \mathbb{R}^n, x \neq 0 \right\} \quad (7.42)$$

(compruébense estas afirmaciones).

Si $T \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ y $S \in \mathcal{L}(\mathbb{R}^m, \mathbb{R}^p)$, entonces la composición ST satisface $ST \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^p)$ y

$$\|ST\| \leq \|S\| \|T\| \quad (7.43)$$

En efecto, basta tener en cuenta que, siendo $x \in \bar{B}_1(0)$,

$$|(ST)(x)| = |S(Tx)| \leq \|S\| |Tx| \leq \|S\| \|T\|$$

y que $\|ST\| = \max\{|(ST)x|; |x| \leq 1\}$. (Como se sabe, el producto de dos matrices puede ser la matriz cero aunque no lo sea ninguno de los factores,

por lo que en (7.43) no se puede remplazar la desigualdad por una igualdad.). En particular, si $L \in \mathcal{L}(\mathbb{R}^n) \approx \mathcal{M}_n \approx \mathbb{R}^{n^2}$, entonces

$$\|L^k\| \leq \|L\|^k \quad k = 0, 1, 2, \dots \quad (7.44)$$

Hay otras normas de matrices (aplicaciones lineales) que también verifican las importantes desigualdades (7.41) y (7.43) y que, frecuentemente, son más fáciles de manejar que la norma uniforme. Así, es fácil comprobar (hágase como ejercicio) que la norma

$$|A|_1 = \sum_{i=1}^m \sum_{j=1}^n |a_{ij}| \quad (7.45)$$

verifica (7.41) cuando se toma en \mathbb{R}^n y \mathbb{R}^m la norma $|\cdot|_1$, así como la desigualdad (7.43) para el producto de dos matrices.

Insistamos en que la norma uniforme depende de las normas utilizadas en los espacios \mathbb{R}^n implicados (problemas 30 y 31). Si no se indica nada en contra, se entenderá que se utiliza en \mathbb{R}^n la norma euclídea.

7.7 Conexión

Como decíamos en el capítulo 4, la idea intuitiva de conjunto conexo es la de que está hecho de una sola pieza. La manera precisa y general de fijar esa idea es mediante la siguiente

Definición 7.18 Se dice que $S \subseteq \mathbb{R}^n$ es **conexo** si **no** es posible encontrar dos conjuntos abiertos A_1 y A_2 tales que

- i) $C_1 = S \cap A_1 \neq \emptyset$, $C_2 = S \cap A_2 \neq \emptyset$
- ii) $C_1 \cap C_2 = \emptyset$
- iii) $S = C_1 \cup C_2$.

En otras palabras, $S \subseteq \mathbb{R}^n$ es conexo si **no** existe ninguna pareja de abiertos relativos a S no vacíos C_1 y C_2 tales que $S = C_1 \cup C_2$ y $C_1 \cap C_2 = \emptyset$, lo que equivale a decir que el único subconjunto no vacío de S que es a la vez abierto y cerrado relativo a S es el propio S (figura 7.3).

Después veremos que la definición que dimos en el capítulo 4 de abierto conexo es equivalente a ésta.

Los intervalos de la recta real \mathbb{R} son los ejemplos más simples que responden a la idea intuitiva de conjunto conexo, y efectivamente se tiene:

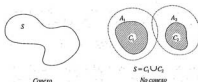


Figura 7.3

Teorema 7.17 *Un subconjunto S de \mathbb{R} es conexo si y sólo si es un intervalo.*

Daremos la demostración al final del apartado. En particular, \mathbb{R} es conexo, es decir, los únicos conjuntos de \mathbb{R} que son a la vez abiertos y cerrados son el propio \mathbb{R} y el subconjunto vacío. Lo mismo ocurre para \mathbb{R}^n , como veremos enseguida (véase también el problema 39).

Las funciones continuas conservan la propiedad de conexión:

Teorema 7.18 *Sea $f : S \rightarrow \mathbb{R}^m$ continua en un conjunto conexo $S \subseteq \mathbb{R}^n$. Entonces $f(S)$ es conexo.*

Demostración: Supongamos que $f(S)$ no es conexo. Así $f(S) = A \cup B$, donde A y B son dos subconjuntos no vacíos de $f(S)$, abiertos relativos a $f(S)$ y tales que $A \cap B = \emptyset$. Se tiene $S = f^{-1}(f(S)) = f^{-1}(A \cup B) =$ (por ser A y B disjuntos) $= f^{-1}(A) \cup f^{-1}(B)$. Pero $f^{-1}(A)$ y $f^{-1}(B)$ son abiertos relativos a S , según vimos al final del apartado 7.4, y disjuntos por serlo A y B . Se concluye entonces que S no es conexo contrariamente a la hipótesis.

Corolario 7.5 *(Teorema de Bolzano o del valor intermedio) Sea $f : S \rightarrow \mathbb{R}$ continua en un conjunto conexo $S \subseteq \mathbb{R}^n$. Entonces $f(S)$ es un intervalo y, en consecuencia, si a y b son dos puntos de $f(S)$ tales que $a < b$, existe, para todo c que verifica $a < c < b$, un $x \in S$ tal que $f(x) = c$.*

Conexión por arcos

Un arco en \mathbb{R}^n es la imagen de una función continua $\varphi : \mathbb{R} \supset [a, b] \rightarrow \mathbb{R}^n$. Se dice que conecta los puntos p y q si $\varphi(a) = p$ y $\varphi(b) = q$. (Con cierto abuso de lenguaje se dice también que la propia función φ es un arco.)

El intervalo $[a, b]$ se puede normalizar al $[0, 1]$, y suponer, si ello interesa, que un arco es la imagen de una función continua $\varphi : [0, 1] \rightarrow \mathbb{R}^n$. En efecto, el arco definido por $\varphi_1 : [0, 1] \rightarrow \mathbb{R}^n$ es el mismo que el definido por la función

$\varphi_2 : [a, b] \rightarrow \mathbb{R}^n$ dada por $\varphi_2 = \varphi_1 \circ \phi$ donde ϕ es la función continua estrictamente creciente

$$\phi(t) = \frac{t-a}{b-a}$$

(cuya función inversa es $s \mapsto a + (b-a)s$).

Observaciones:

1. Para todo $p \in \mathbb{R}^n$, la función continua $\varphi(t) = p$ para todo $t \in [0, 1]$ define un arco constante en p .
2. Si $\varphi : [0, 1] \rightarrow \mathbb{R}^n$ es un arco que conecta p y q , entonces $\tilde{\varphi} : [0, 1] \rightarrow \mathbb{R}^n$ dada por $\tilde{\varphi}(t) = \varphi(1-t)$ es un arco que conecta q con p .
3. Sean $\varphi_k : [0, 1] \rightarrow \mathbb{R}^n$, $k = 1, \dots, m$, arcos con la propiedad de que $\varphi_k(1) = \varphi_{k+1}(0)$, $k = 1, \dots, m-1$ (o sea, el punto final p_k del arco definido por φ_k es el punto inicial del definido por φ_{k+1}). Entonces el arco

$$\varphi(t) = \varphi_k(t - (k-1)) \text{ para } k-1 \leq t \leq k, k = 1, \dots, m$$

definido en el intervalo $t \in [0, m]$ es la yuxtaposición de los arcos $\varphi_1, \dots, \varphi_m$ y conecta $p_0 = \varphi_1(0)$ con $p_m = \varphi_m(1)$ pasando por los puntos p_1, \dots, p_{m-1} . En particular, una poligonal es una yuxtaposición de segmentos.

4. Un arco es un conjunto conexo por ser la imagen continua de un intervalo.

Definición 7.19 Un conjunto $S \subseteq \mathbb{R}^n$ es *conexo por arcos* si dos puntos cualesquiera de S se pueden conectar por un arco completamente contenido en S .

Por ejemplo, un conjunto convexo es conexo por arcos ya que para todo par de puntos p y q , el segmento que los une, imagen de la función continua

$$[0, 1] \ni t \mapsto p + t(q-p)$$

está, por definición de convexidad, completamente contenido en el conjunto.

Teorema 7.19 Un conjunto $S \subseteq \mathbb{R}^n$ conexo por arcos es conexo.

Demostración: Si S no es conexo, existen dos abiertos A_1 y A_2 de \mathbb{R}^n tales que $C_1 = A_1 \cap S$, $C_2 = A_2 \cap S$ son conjuntos no vacíos disjuntos cuya unión es S . Sea $p \in C_1$ y $q \in C_2$; por hipótesis, hay una función $\varphi : [0, 1] \rightarrow S$ tal que $\varphi(0) = p$, $\varphi(1) = q$. Por ser φ continua, $\varphi^{-1}(C_1)$ y $\varphi^{-1}(C_2)$ son abiertos relativos a $[0, 1]$, no vacíos ($0 \in \varphi^{-1}(C_1)$, $1 \in \varphi^{-1}(C_2)$) y tales que su intersección es vacía y su unión todo $[0, 1]$; pero esto contradice la conexión del intervalo $[0, 1]$. S ha de ser, por tanto, conexo.

La conexión por arcos es una propiedad muy intuitiva que permite, gracias a este teorema, asegurar con cierta facilidad la conexión de muchos conjuntos (por ejemplo, el espacio total \mathbb{R}^n , que es convexo, es conexo, y lo mismo las bolas abiertas y las bolas cerradas). Pero, aunque resulte un tanto chocante a la intuición, el recíproco del teorema no es cierto, es decir, hay conjuntos conexos que no son conexos por arcos. Un contraejemplo bien conocido es el conjunto $S = A \cup B$ donde

$$\begin{aligned} A &= \{(x, \sin(1/x)) ; 0 < x \leq 1\} \\ B &= \{(0, y) ; -1 \leq y \leq 1\} \end{aligned}$$

(figura 7.4). A es conexo por ser imagen continua de un intervalo y $\bar{A} = S$, con lo que, por el problema 40, también es conexo. Sin embargo, S no es conexo por arcos, pues no se puede conectar mediante un arco un punto de A con un punto de B .

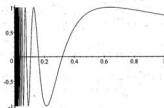


Figura 7.4

Las dos nociones de conexión coinciden, no obstante, en conjuntos abiertos:

Teorema 7.20 Si $S \subseteq \mathbb{R}^n$ es un conjunto no vacío abierto y conexo entonces es conexo por arcos.

Demostración: Sea $p \in S$ y sea A el subconjunto de puntos de S que se pueden conectar a p mediante un arco y $B = \{x \in S ; x \notin A\}$. Entonces

S es la unión de dos conjuntos disjuntos. Se trata de probar que ambos son abiertos, con lo que, al ser $A \neq \emptyset$, ya que $p \in A$, y S conexo por hipótesis, habrá de ser necesariamente $B = \emptyset$ y, en consecuencia, $S = A$. Comencemos por A . Sea $q \in A$; como S es abierto, existe una bola $B_\epsilon(q) \subset S$; $B_\epsilon(q)$ es conexo por ser conexo por arcos (es un conjunto convexo); por tanto, todo $z \in B_\epsilon(q)$ se puede conectar con q , que, por hipótesis, se puede conectar con p ; yuxtaponiendo los dos arcos, podemos conectar cualquier $z \in B_\epsilon(q)$ con p ; así pues, $B_\epsilon(q) \subset A$ y, en consecuencia, A es abierto.

Veamos ahora que B también es abierto. Sea $x \in B$. Por ser S abierto, existe una bola $B_\delta(x) \subset S$. Si algún $z \in B_\delta(x)$ se pudiese conectar mediante un arco a p , lo mismo podría hacerse con x , ya que, como antes, x se puede conectar con p por ser $B_\delta(x)$ conexo por arcos. Como eso no puede ser, pues $x \notin A$, se tendrá $B_\delta(x) \subset B$ y B es abierto. Así pues, se tiene, como decíamos, $S = A$. Pero A es conexo por arcos pues dados dos puntos cualesquiera $x, y \in A$, se pueden conectar yuxtaponiendo un arco que vaya de x a p con uno que vaya de p a y .

De hecho, los puntos de un conjunto abierto y conexo se pueden unir por poligonales, que son un caso particular de arcos, y, así, la definición de abierto conexo que dimos en el capítulo 4 no suponía ninguna pérdida de generalidad:

Teorema 7.21 *Si D es un conjunto abierto y conexo, dos cualesquiera de sus puntos se pueden unir mediante una poligonal completamente contenida en D .*

Demostración: En realidad, vale la demostración anterior llamando A al conjunto de puntos que se pueden conectar con p mediante una poligonal, pues obsérvese en los argumentos utilizados que, por ser $B_\epsilon(q)$ y $B_\delta(x)$ conjuntos convexos, lo que se añade a una poligonal se puede elegir que sea un segmento para obtener otra poligonal (y también se puede aplicar la misma idea para demostrar que el enunciado de este teorema sigue siendo válido si se sustituye "poligonal" por "poligonal de elementos paralelos a los ejes"; hágase como ejercicio). Tiene cierto interés otra demostración basada en la aproximación de un arco cualquiera por una poligonal: sean p, q dos puntos de D . Por ser D conexo por arcos, existe un arco $\varphi: [0, 1] \rightarrow D$ que los une. $\varphi([0, 1])$ es un subconjunto compacto de D , por lo que su distancia a la frontera de D será un número positivo ϵ (problema 17). Por la continuidad uniforme de φ , existe $\delta > 0$ tal que

$$t, s \in [0, 1], |t - s| < \delta \implies |\varphi(t) - \varphi(s)| < \epsilon \quad (7.46)$$

Sea m un número natural tal que

$$\frac{1}{m} \leq \delta$$

Con el paso $h = 1/m$ se obtiene una partición del intervalo $[0, 1]$

$$0 = t_0 < t_1 = h < \dots < t_k = kh < \dots < t_m = mh = 1$$

que permite formar la poligonal $[\varphi(t_0), \varphi(t_1), \dots, \varphi(t_m)]$ que une $p = \varphi(t_0)$ con $q = \varphi(t_m)$ y que, por (7.46), está completamente contenida en D . (Ejercicio: Hágase una tercera demostración del teorema utilizando el teorema de Heine-Borel.)

Demostración del teorema 7.17

Recordemos que hay diez tipos de intervalos en \mathbb{R} :

$$\begin{aligned} &[a, a], (a, b), (a, b], [a, b], [a, b], \text{ intervalos finitos} \\ &(-\infty, b), (-\infty, b], (a, \infty), [a, \infty), (-\infty, \infty), \text{ intervalos infinitos} \end{aligned}$$

Aparte de los constituidos por un solo punto, los restantes nueve tipos de intervalos I están caracterizados por la siguiente propiedad

$$a, b \in I, a < x < b \Rightarrow x \in I \quad (7.47)$$

En efecto, supongamos que I es un subconjunto no vacío de \mathbb{R} que verifica esta propiedad y que $c \in I$; pongamos $p = \inf I$ y $q = \sup I$. Si $p = -\infty$, entonces, para todo $x < c$ existe $y \in I$ tal que $y < x$, con lo que, por (7.47), $x \in I$ y $(-\infty, c] \subseteq I$. Si p es finito y $p < c$, para todo x que satisface $p < x < c$ existe $y \in I$ tal que $p < y < x$, con lo que, de nuevo por (7.47), $x \in I$ y también $(p, c] \subseteq I$. De la misma manera se comprueba que $I \supseteq [c, q)$ si $q > c$. Se tiene entonces que si $p < c < q$, $I \supseteq (p, q) = (\inf I, \sup I)$ y como, naturalmente, $I \subseteq (\inf I, \sup I)$, resulta que I ha de ser un intervalo de uno de los tipos indicados:

$$\begin{aligned} p > -\infty, p \notin I, q < \infty, q \notin I &\Rightarrow I = (p, q) \\ p > -\infty, p \notin I, q < \infty, q \in I &\Rightarrow I = (p, q] \\ &\text{etc.} \end{aligned}$$

Veamos la demostración del teorema. Sea S un conjunto conexo de \mathbb{R} . Si S consta de un solo punto (los conjuntos de un solo punto son conexos), es un intervalo; supongamos entonces que S contiene dos puntos distintos $a < b$; se trata de probar que todo x tal que $a < x < b$ pertenece a S , de lo que se deducirá por la caracterización anterior que S es un intervalo. Si no fuese así, o sea, si existiese un $\bar{x} \notin S$ tal que $a < \bar{x} < b$, entonces se tendría

$$S = \{S \cap (-\infty, \bar{x})\} \cup \{S \cap (\bar{x}, \infty)\}$$

y S no sería conexo contra la hipótesis.

Recíprocamente, supongamos que S es un intervalo. Si S no es conexo, $S = C_1 \cup C_2$, donde C_1 y C_2 son dos conjuntos abiertos relativos a S no vacíos tales que $C_1 \cap C_2 = \emptyset$. Por ser conjuntos no vacíos, existen $x \in C_1$ e $y \in C_2$ tales que $x \neq y$; se puede suponer, cambiando la notación si es necesario, $x < y$. Puesto que S es un intervalo, $[x, y] \subseteq S$ y cada punto de $[x, y]$ pertenece bien a C_1 bien a C_2 . Sea $a = \sup \{C_1 \cap [x, y]\}$ (que es finito por tratarse de un conjunto acotado y pertenece a $[x, y]$ que es cerrado); Si $a \in C_1$, entonces $a < y$, y existe un intervalo $[a, a+h]$ contenido en $[x, y] \subseteq S$ y en C_1 (por ser éste abierto relativo a S), lo que contradice la definición de a como supremo. Si, alternativamente, $a \in C_2$, entonces $x < a$ y, como antes, existe un entorno $(a-h, a] \subset C_2 \cap [x, y]$, lo que también contradice la definición de a como supremo, pues habría todo un entorno de puntos a la izquierda de a que no pertenecerían al conjunto $C_1 \cap [x, y]$ del que se supone es supremo. Resulta, por tanto, que a no pertenece ni a C_1 ni a C_2 lo que es absurdo pues, como veíamos, $a \in [x, y] \subseteq S$ y $S = C_1 \cup C_2$. Así pues, S ha de ser conexo.

7.8 Notas y complementos

Los distintos grupos de propiedades del espacio \mathbb{R}^n sirven como modelo para definir estructuras matemáticas más generales que \mathbb{R}^n que, en diversos contextos, han resultado ser el marco apropiado para el planteamiento y resolución de importantes problemas. Por ejemplo, en la rama del análisis matemático llamada *análisis funcional* se estudian conjuntos de funciones, en lugar de funciones aisladas, como espacios abstractos con determinada estructura en los que sus elementos o "puntos" son las propias funciones, y ello con la idea, en particular, de resolver *ecuaciones funcionales* tales como ecuaciones diferenciales, ecuaciones integrales y otras.

En este apartado vamos a comentar brevemente, de mayor a menor generalidad, las características de las estructuras más importantes que, sobre el modelo de \mathbb{R}^n , tratan de dar un sentido matemático a las ideas intuitivas de *proximidad*, *convergencia*, *continuidad*, *aproximación*,...

1. Espacios topológicos

Sea X un conjunto no vacío. Se dice que una familia \mathcal{T} de subconjuntos de X es una **topología** en X o que define una **estructura topológica** en X si \mathcal{T} satisface las propiedades de la proposición 7.1:

- (a) X y el conjunto vacío \emptyset pertenecen a \mathcal{T} .
- (b) La unión de conjuntos de \mathcal{T} es un conjunto de \mathcal{T} .

- (c) La intersección de una colección finita de conjuntos de \mathcal{T} es un conjunto de \mathcal{T} .

Los conjuntos de \mathcal{T} son los conjuntos *abiertos* para dicha topología. X junto con \mathcal{T} , o sea, el par (X, \mathcal{T}) se dice que es un **espacio topológico**. Los elementos de X se llaman *puntos*. Sobre un mismo conjunto X *subyacente* pueden estar definidas distintas topologías y tenerse, en consecuencia, distintos espacios topológicos. Si no hay riesgo de confusión, se puede nombrar al espacio topológico simplemente por X .

La familia de conjuntos abiertos que hemos definido en \mathbb{R}^n es la topología usual. Para cualquier conjunto X no vacío, la familia $\mathcal{P}(X)$ de todos los subconjuntos de X satisface los axiomas (a)(b)(c) de las estructuras topológicas; se la denomina topología *discreta* sobre X . En el extremo opuesto, la topología *indiscreta* sobre X es la formada únicamente por X y \emptyset .

Sea A un subconjunto no vacío de un espacio topológico (X, \mathcal{T}) . Se comprueba fácilmente que la familia de intersecciones de A con los conjuntos de \mathcal{T} constituye una topología en A . Es la *topología relativa* o *inducida* en A por la topología \mathcal{T} de X . Con ella, se dice que A es un *subespacio* de X .

Obsérvese que, por ejemplo, en la topología inducida en el intervalo cerrado $[1, 4]$ por la topología usual de \mathbb{R} , el intervalo $[1, 3]$ es abierto pues es la intersección de $[1, 4]$ con el intervalo abierto $(0, 3)$ de \mathbb{R} .

Un *entorno* de un punto $x \in X$ es un conjunto que contiene a un conjunto abierto que contiene a x . También se puede definir una topología en términos del sistema de entornos de un punto. Un conjunto es abierto si y sólo si es entorno de todos sus puntos.

Un conjunto *cerrado* es el complementario en X de un conjunto abierto. Los conjuntos cerrados satisfacen las propiedades de la proposición 7.3.

Las definiciones de *punto interior*, *punto frontera*, *punto exterior* y *punto de acumulación* son las mismas que hemos visto en \mathbb{R}^n utilizando la definición anterior de entorno.

Sean (X, \mathcal{T}) y (X', \mathcal{T}') dos espacios topológicos y sea $f: X \rightarrow X'$. Se dice que f es *continua* en un punto $x_0 \in X$ si para todo entorno V' de $f(x_0)$ en X' existe un entorno V de x_0 en X tal que $x \in V$ implica $f(x) \in V'$. Se dice que $f: X \rightarrow X'$ es continua en X (o, simplemente, que es continua) si es continua en todo punto de X . No es difícil probar el siguiente

Teorema 7.22 Sean (X, \mathcal{T}) y (X', \mathcal{T}') dos espacios topológicos y sea $f: X \rightarrow X'$. Las tres afirmaciones siguientes son equivalentes:

- (i) f es continua en X ,
- (ii) Para todo conjunto abierto A' de X' , $f^{-1}(A')$ es un conjunto abierto de X ,

- (iii) Para todo conjunto cerrado F' de X' , $f^{-1}(F')$ es un conjunto cerrado de X .

Si $f : X \rightarrow X'$ es biyectiva y tanto f como f^{-1} son continuas, se dice que f es un *homeomorfismo* y que los espacios X y X' son *homeomorfos*.

Una sucesión $\{x_k\}$ de puntos de un espacio topológico (X, T) se dice que converge a un límite $a \in X$ si para todo entorno A de a existe un número entero N tal que $x_k \in A$ para todo $k \geq N$. El límite de una sucesión puede no ser único (por ejemplo, en un conjunto X con la topología indiscreta cualquier sucesión tiene por límite todo punto de X). En los espacios métricos que consideraremos a continuación, el límite de una sucesión convergente es único. Las funciones continuas conservan la convergencia de sucesiones, es decir, si $f : X \rightarrow X'$ es continua en x_0 , entonces, para toda sucesión $\{x_k\} \subset X$ tal que $x_k \rightarrow x_0$ se tiene que $f(x_k) \rightarrow f(x_0)$ en la topología T' de X' . Pero, a diferencia de lo que vimos en \mathbb{R}^n , el recíproco no es cierto, es decir, puede ocurrir en un espacio topológico general que una función conserve la convergencia de sucesiones (y se dice entonces que es *continua por sucesiones*) y que, sin embargo, no sea continua. Es necesario introducir en los espacios topológicos objetos más generales que las sucesiones para que la continuidad pueda caracterizarse en términos de convergencia.

Un subconjunto K de un espacio topológico X es *compacto* si de todo recubrimiento abierto de K se puede extraer un recubrimiento finito. (Las definiciones de recubrimiento y recubrimiento abierto son las mismas que en 7.3). Si X , considerado como subconjunto de sí mismo, es compacto, se dice que es un *espacio topológico compacto*. No es difícil probar que si $f : X \supseteq K \rightarrow X'$ es continua en el conjunto compacto K , entonces $f(K)$ es un conjunto compacto.

Un conjunto A es *compacto por sucesiones* si de toda sucesión $\{x_k\}$ de elementos de A se puede extraer una subsucesión que converge a un punto de A . En general, existen conjuntos compactos que no son compactos por sucesiones y viceversa, aunque en los espacios métricos que vamos a ver a continuación ambas nociones coinciden (lo que sí implican ambas nociones es la de *numerablemente compacto*, propiedad que tiene un conjunto A cuando de todo recubrimiento abierto numerable de A se puede extraer un recubrimiento finito, y que, a su vez, es equivalente a la propiedad de que toda sucesión de elementos de A posee un punto de acumulación que pertenece a A . Para caracterizar la compacidad en términos de convergencia o de puntos de acumulación se necesitan, como en el caso de la continuidad, conceptos más complejos.)

La definición de conjunto *conexo* es la misma que hemos dado en \mathbb{R}^n .

2. Espacios métricos

Como decíamos en el apartado 1.4, una *distancia* o *métrica* en un conjunto X es una aplicación $d: X \times X \rightarrow [0, \infty)$ que satisface las propiedades:

$$d(x, y) \geq 0 \text{ y } d(x, y) = 0 \text{ si y sólo si } x = y$$

$$d(x, y) = d(y, x)$$

$$d(x, z) \leq d(x, y) + d(y, z) \text{ (desigualdad triangular)}$$

Un conjunto X dotado de una métrica d , o, con más precisión, el par (X, d) , se dice que es un *espacio métrico*. El modelo de espacio métrico es, naturalmente, el espacio \mathbb{R}^n con la *métrica euclídea*

$$d(x, y) = |x - y| = \sqrt{(x_1 - y_1)^2 + \dots + (x_n - y_n)^2}$$

En cualquier conjunto no vacío X , la función

$$d(x, y) = \begin{cases} 0 & \text{si } x = y \\ 1 & \text{si } x \neq y \end{cases}$$

define una métrica llamada *métrica trivial* o *discreta*.

En el conjunto de sucesiones de números reales $x = (x_1, x_2, \dots, x_i, \dots)$, denotado por \mathbb{R}^∞ , consideremos el subconjunto l_∞ formado por las sucesiones para las que

$$\sup_{i \in \mathbb{N}} |x_i| < \infty$$

Se comprueba sin dificultad que la función

$$d(x, y) = \sup_{i \in \mathbb{N}} |x_i - y_i|$$

es una métrica en l_∞ .

La mayoría de los conceptos que hemos introducido en \mathbb{R}^n en este capítulo se definen de la misma manera en cualquier espacio métrico: la *bola abierta* de centro x_0 y radio ϵ es el conjunto $B_\epsilon(x_0) = \{x \in X; d(x, x_0) < \epsilon\}$. Con ellas se definen como en \mathbb{R}^n los entornos, los conjuntos abiertos y cerrados, etc. La familia \mathcal{T} de conjuntos abiertos así definidos constituye la topología en X inducida por la métrica d . A su vez, un espacio topológico cuya topología esté generada por una métrica se dice que es *metrizable*.

Si A es un subconjunto no vacío de X , la restricción de la distancia d a los elementos de A define obviamente una distancia d_A en A , y, con ella, A

es un subespacio (métrico) de X cuya topología asociada es la inducida por la topología \mathcal{T} de X .

Una sucesión $\{x_k\}$ de puntos de X converge a un punto $x \in X$ si $d(x_k, x) \rightarrow 0$, o, equivalentemente, si para todo entorno U de x existe $N \in \mathbb{N}$ tal que $x_k \in U$ para todo $k \geq N$. El límite de una sucesión es único.

Un conjunto $A \subset X$ es *acotado* si existe M tal que $d(x, x_0) \leq M$ para todo $x \in A$, siendo x_0 un elemento fijo de A . Los puntos de una sucesión convergente forman un conjunto acotado.

Una sucesión $\{x_k\}$ de puntos de X es de Cauchy si para todo $\varepsilon > 0$ existe $N \in \mathbb{N}$ tal que $d(x_p, x_q) < \varepsilon$ para todo $p, q \geq N$. Como en \mathbb{R}^n , toda sucesión de Cauchy es acotada y toda sucesión convergente es de Cauchy, pero el recíproco de esto último no es cierto en general. Un *espacio métrico completo* es aquél en el que toda sucesión de Cauchy es convergente. \mathbb{R}^n es un espacio métrico completo y se comprueba fácilmente que $A \subset \mathbb{R}^n$ es completo si y sólo si es cerrado. (Por ejemplo, el intervalo $(0, 1)$ con la métrica usual inducida por la de \mathbb{R} , no es completo pues, por ejemplo, la sucesión $(\frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \dots)$ es de Cauchy pero no converge a un punto de $(0, 1)$.)

También como en \mathbb{R}^n , se comprueba que los conceptos de conjunto compacto y compacto por sucesiones son equivalentes. También se cumple que todo conjunto compacto de un espacio métrico es cerrado y acotado. Sin embargo, y esto es una importante diferencia, el recíproco no es cierto: consideremos en l_∞ el conjunto $\bar{B}_1(0) = \{x; d(x, 0) \leq 1\}$, donde 0 indica la sucesión nula $(0, \dots, 0, \dots)$; es un conjunto cerrado y acotado; la sucesión $\{e_i\}$, $i \in \mathbb{N}$, de elementos de $\bar{B}_1(0)$, donde e_i es la sucesión de números reales con un 1 en el lugar i y ceros en los demás, no posee ninguna subsucesión convergente ya que $d(e_i, e_j) = 1$ para todo i, j con $i \neq j$. (Para la validez del resultado en \mathbb{R}^n es decisiva la dimensión finita del espacio.)

3. Espacios normados

Las estructuras topológicas y métricas se definen sobre un conjunto cualquiera. Ahora, se parte ya de que el conjunto subyacente está dotado de la estructura algebraica de espacio vectorial (real, para simplificar, aunque también pueden considerarse espacios sobre el cuerpo de los números complejos), la cual, en un sentido muy preciso, va a ser compatible con la estructura topológica que se va a introducir en él. Una norma en un espacio vectorial E es, como vimos en el apartado 1.4, una aplicación

$$N : E \longrightarrow [0, \infty)$$

que satisface las propiedades

$$N(x) \geq 0 \text{ para todo } x \in E \text{ y } N(x) = 0 \text{ si y sólo si } x = 0$$

$$N(x+y) \leq N(x) + N(y) \text{ (desigualdad triangular)}$$

$$N(cx) = |c| N(x), \quad c \in \mathbb{R}$$

E , dotado de una norma, se dice que es un *espacio normado*.

A una norma $N : E \rightarrow [0, \infty)$ está asociada una **distancia** entre puntos de E dada por

$$d_N(x, y) = N(x - y)$$

como se comprueba fácilmente. Así pues, un espacio normado es un tipo especial de espacio métrico, y, por tanto, de espacio topológico. La compatibilidad aludida entre las estructuras de espacio vectorial y de espacio topológico se expresa en que las aplicaciones

$$E \times E \rightarrow \mathbb{R} \text{ dada por } (x, y) \rightarrow x + y \text{ (suma de vectores)}$$

$$\mathbb{R} \times E \rightarrow \mathbb{R} \text{ dada por } (c, x) \rightarrow cx \text{ (ley externa)}$$

son continuas (para dar sentido a esta afirmación hay que dotar a los productos cartesianos $E \times E$ y $\mathbb{R} \times E$ de las normas $N((x, y)) = N(x) + N(y)$ y $N((c, x)) = |c| + N(x)$ respectivamente. Cuando en un espacio vectorial se tiene definida una topología que es compatible en ese sentido con la estructura de espacio vectorial, se tiene un *espacio vectorial topológico*, estructura que generaliza la de espacio normado.)

Como vimos en el apartado 7.5, \mathbb{R}^n se puede dotar de diversas estructuras de espacio normado, si bien todas ellas dan lugar a la misma topología. Esto se debía a la equivalencia de todas las normas que se pueden definir en \mathbb{R}^n , resultado que se basaba en la compacidad de la bola unidad. Esto último no tiene por que ser cierto en un espacio normado general; veamos: el conjunto \mathbb{R}^∞ de todas las sucesiones de números reales es un espacio vectorial con las operaciones (generalización de las correspondientes operaciones en \mathbb{R}^n):

$$x + y = (x_1 + y_1, x_2 + y_2, \dots)$$

$$cx = (cx_1, cx_2, \dots)$$

l_∞ es un subespacio vectorial de \mathbb{R}^∞ , es un espacio normado con la norma

$$|x|_\infty = \sup_{i \in \mathbb{N}} |x_i|$$

y acabamos de ver que, en él, la bola unidad no es compacta. Ocurre entonces que, en general, dos normas distintas en un espacio vectorial no son necesariamente equivalentes según la definición 7.17, y, en consecuencia, no definen necesariamente la misma topología. Esta afirmación se refiere, en realidad, a espacios vectoriales de dimensión infinita como l_∞ , pues se puede demostrar que si un espacio normado es tal que la bola unidad cerrada $\{x; N(x) \leq 1\}$ es compacta, entonces es de dimensión finita.

Un espacio normado completo se llama *espacio de Banach*.

Otros ejemplos de espacios normados

1. En \mathbb{R}^∞ se considera, para $p \in [1, \infty)$, el conjunto l_p de sucesiones tales que

$$\sum_{i=1}^{\infty} |x_i|^p < \infty$$

Las desigualdades de Hölder y Minkowski del problema 24 se extienden sin dificultad a series mediante un paso al límite y permiten demostrar que l_p es un espacio normado (un espacio de Banach, de hecho) para la norma

$$|x| = \left(\sum_{i=1}^{\infty} |x_i|^p \right)^{1/p}$$

2. Dado un intervalo $I = [a, b]$ cerrado y acotado, llamamos $C = C(I, \mathbb{R})$ al conjunto de funciones continuas reales definidas en I . En C están definidas de manera natural las operaciones de suma y producto por números reales, con respecto a las cuales tiene estructura de espacio vectorial (de dimensión infinita, pues contiene, en particular, a todos los polinomios). Definimos en C la *norma del supremo*

$$\|\varphi\|_0 = \sup\{|\varphi(t)|; t \in I\}$$

bien definida para cada $\varphi \in C$ por el teorema de Weierstrass. Se comprueba fácilmente que, en efecto, $(C, \|\cdot\|_0)$ es un espacio normado utilizando propiedades elementales del supremo (aquí, máximo, de hecho). La convergencia asociada a esta norma $\|\cdot\|_0$ es la *convergencia uniforme* que hemos visto en el apartado 7.4; vimos allí que si una sucesión $\{\varphi_k\}$ de funciones continuas tiene la propiedad de Cauchy uniforme, o sea, es una sucesión de Cauchy para esta norma, entonces converge

—uniformemente— a una función φ continua, lo que traducido al contexto presente significa que C es completo, es decir, un espacio de Banach. Lo anterior vale también para el espacio $C(K, \mathbb{R}^m)$ de funciones continuas de un compacto $K \subset \mathbb{R}^n$ con valores en \mathbb{R}^m con la norma $\|\varphi\|_0 = \sup\{|\varphi(t)|; t \in K\}$ (no importa qué normas se consideren en \mathbb{R}^n y \mathbb{R}^m pues todas son equivalentes).

3. En $C([a, b], \mathbb{R})$ también se puede considerar la familia de normas

$$\|\varphi\|_p = \left(\int_a^b |\varphi(t)|^p dt \right)^{1/p}$$

con $p \in [1, \infty)$. (Compruébese, para $p = 1$, que se trata efectivamente de una norma; se tienen espacios normados que no son de Banach.)

4. Espacios euclídeos

La última estructura de \mathbb{R}^n cuya generalización vamos a comentar es la que deriva del producto escalar. Dado un espacio vectorial (real) E , toda aplicación

$$E \times E \ni (x, y) \mapsto \langle x, y \rangle \in \mathbb{R}$$

que satisfaga las siguientes propiedades:

$$\langle x, y \rangle = \langle y, x \rangle \quad (\text{simetría})$$

$$\langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle; \quad \langle cx, y \rangle = c\langle x, y \rangle, \quad c \in \mathbb{R} \quad (\text{bilinealidad})$$

$$\langle x, x \rangle \geq 0 \quad \text{y} \quad \langle x, x \rangle = 0 \quad \text{si y sólo si} \quad x = 0 \quad (\text{positividad})$$

se denomina producto escalar o interior en E y E dotado de un producto escalar se dice que es un *espacio euclídeo*. Un *espacio de Hilbert* es un espacio euclídeo completo. Los espacios de Hilbert representan la generalización más inmediata de \mathbb{R}^n —muchas de sus propiedades son el trasunto de resultados bien conocidos de la geometría euclídea clásica— y son muy importantes en el análisis funcional y sus aplicaciones. Por ejemplo, en un espacio euclídeo se dispone de la noción de *ortogonalidad* estableciendo que dos vectores son ortogonales si su producto escalar es nulo.

Un resultado importante en cualquier espacio euclídeo E es la *desigualdad de Cauchy-Schwarz*:

$$|\langle x, y \rangle|^2 \leq \langle x, x \rangle \langle y, y \rangle$$

para todo par de vectores $x, y \in E$. La demostración es la misma que vimos en el apartado 1.4 para \mathbb{R}^n , pues sólo utiliza las propiedades que definen al producto escalar. Con él se prueba que

$$\langle x + y, x + y \rangle^{1/2} \leq \langle x, x \rangle^{1/2} + \langle y, y \rangle^{1/2}$$

y, como consecuencia, que

$$E \ni x \mapsto |x| \stackrel{\text{def}}{=} \langle x, x \rangle^{1/2}$$

es una norma en E . Así pues, un espacio euclídeo es un caso particular de espacio normado y un espacio de Hilbert, de espacio de Banach.

En todo espacio euclídeo se satisface la *ley del paralelogramo*

$$|x + y|^2 + |x - y|^2 = 2(|x|^2 + |y|^2)$$

y, recíprocamente, si en un espacio normado E la norma satisface esta ley, la aplicación

$$(x, y) \mapsto \langle x, y \rangle = \frac{1}{4}(|x + y|^2 - |x - y|^2)$$

define un producto escalar cuya norma asociada $\langle x, x \rangle^{1/2}$ es la norma original de E (problema 26).

El espacio l_2 de sucesiones de \mathbb{R}^∞ tales que

$$\sum_{i=1}^{\infty} |x_i|^2 < \infty$$

es un espacio de Hilbert con el producto escalar, generalización natural del producto escalar en \mathbb{R}^n ,

$$\langle x, y \rangle = \sum_{i=1}^{\infty} x_i y_i$$

(casi se podría decir que l_2 es "el" espacio de Hilbert, pues se puede demostrar que todo espacio de Hilbert separable es isomorfo a l_2 ; un espacio métrico se dice que es *separable* si posee un subconjunto a lo más numerable que es denso en el espacio; por ejemplo, \mathbb{R} es separable pues el conjunto \mathbb{Q} de números racionales es numerable y denso en \mathbb{R}).

Otro espacio euclídeo interesante es $C([a, b], \mathbb{R})$ con el producto escalar

$$\langle f, g \rangle = \int_a^b f(t)g(t)dt$$

que da lugar a la norma

$$\|f\|_2 = \left(\int_a^b |f(t)|^2 dt \right)^{1/2}$$

No es un espacio completo (se puede incluir como subespacio denso en un espacio completo, es decir, un espacio de Hilbert, denotado por $L^2[a, b]$, que juega un papel fundamental en la teoría de integración de Lebesgue).

Compruébese como ejercicio que el espacio $C = C([a, b], \mathbb{R})$ con la norma

$$\|\varphi\|_0 = \sup\{|\varphi(t)|; t \in [a, b]\}$$

no es espacio de Hilbert (**Indicación:** Tómese, por ejemplo, $[a, b] = [0, 3]$; sea φ igual a 1 en $[0, 1]$, a 0 en $[2, 3]$ y lineal en $[1, 2]$ y, análogamente, sea ψ igual a 0 en $[0, 1]$, a 1 en $[2, 3]$ y lineal en $[1, 2]$, de modo que sean funciones continuas. Compruébese que no se verifica para ellas la ley del paralelogramo.)

7.9 Problemas

- (a) Probar que toda intersección finita de entornos de un punto es también entorno del punto.
- (b) Probar que los entornos *separan puntos*, es decir, que si $x \neq y$, existe un entorno $U(x)$ de x y un entorno $U(y)$ de y tales que $U(x) \cap U(y) = \emptyset$.

- (a) Demostrar la proposición 7.1. **Indicación:** Para una buena expresión de los argumentos, recordemos que, si representamos por \mathfrak{F} una colección de conjuntos de \mathbb{R}^n , la *unión* de todos los conjuntos que pertenecen a \mathfrak{F} , es decir, el conjunto de puntos de \mathbb{R}^n que pertenecen *al menos* a uno de los conjuntos de la colección \mathfrak{F} , se denota por $\bigcup\{A; A \in \mathfrak{F}\}$ o también por $\bigcup_{A \in \mathfrak{F}} A$. Si la colección \mathfrak{F} es

finita o numerable se utilizan las notaciones $\bigcup_{j=1}^m A_j = A_1 \cup \dots \cup A_m$

y $\bigcup_{j=1}^{\infty} A_j$. Análogamente, la *intersección* de los conjuntos de \mathfrak{F} , o sea, el conjunto de puntos de \mathbb{R}^n que pertenecen a *todos* los conjuntos de la colección \mathfrak{F} , se denota por $\bigcap\{A; A \in \mathfrak{F}\}$ o $\bigcap_{A \in \mathfrak{F}} A$ y en

el caso finito o numerable por $\bigcap_{j=1}^m A_j = A_1 \cap \dots \cap A_m$ y $\bigcap_{j=1}^{\infty} A_j$.

(b) En relación a la afirmación (c) de la proposición, examinar la intersección $\bigcap_{j=1}^{\infty} A_j$ de los conjuntos $A_j = \{x \in \mathbb{R}^2; |x| < 1 + 1/j\}$.

3. Demostrar que un conjunto es cerrado si y sólo si su complementario es abierto.
4. (a) Demostrar que la bola cerrada $\bar{B}_\varepsilon(x_0)$ es un conjunto cerrado comprobando que su complementario es abierto (aquí hay que utilizar la propiedad de la norma

$$||x| - |y|| \leq |x - y|$$

que se deriva de la propiedad triangular escribiendo

$$\begin{aligned} |x| &= |(x - y) + y| \leq |x - y| + |y| \\ |y| &= |(y - x) + x| \leq |x - y| + |x| \end{aligned}$$

e imitar la demostración del tema 2 de que $B_\varepsilon(x_0)$ es abierto).

- (b) Probar que $\partial B_\varepsilon(x_0) = \partial \bar{B}_\varepsilon(x_0) = \{x \in \mathbb{R}^n; |x - x_0| = \varepsilon\} \stackrel{\text{def}}{=} S_\varepsilon(x_0)$, la esfera $(n-1)$ -dimensional de centro x_0 y radio ε , y que, por tanto, $\bar{B}_\varepsilon(x_0)$ es la adherencia de $B_\varepsilon(x_0)$. (Indicación: la inclusión $\partial B_\varepsilon(x_0) \subset S_\varepsilon(x_0)$ es fácil, ya que si x es tal que $|x - x_0| < \varepsilon$ o $|x - x_0| > \varepsilon$, entonces no puede pertenecer a $\partial B_\varepsilon(x_0)$; para probar que, recíprocamente, todo punto x tal que $|x - x_0| = \varepsilon$ pertenece a $\partial B_\varepsilon(x_0)$, considérese la semirrecta $\{x_0 + \lambda x; \lambda > 0\}$ y compruébese que sobre ella hay puntos de $B_\varepsilon(x_0)$ y también puntos de $(B_\varepsilon(x_0))^c$ todo lo próximos a x que se desee; la demostración para $\partial \bar{B}_\varepsilon(x_0)$ es análoga.)

5. (a) Demostrar la proposición 7.3. Indicación: Hay que tomar complementarios en las propiedades (b) y (c) de la proposición 7.1 y utilizar las *leyes de Morgan* relativas a la interacción entre uniones, intersecciones y complementarios:

Si \mathfrak{F} es una colección de conjuntos cualquiera se tiene

$$\left(\bigcup \{A; A \in \mathfrak{F}\} \right)^c = \bigcap \{A^c; A \in \mathfrak{F}\}$$

$$\left(\bigcap \{A; A \in \mathfrak{F}\} \right)^c = \bigcup \{A^c; A \in \mathfrak{F}\}$$

- (b) En relación a la afirmación (c) de la proposición, examinar la unión $\bigcup_{j=1}^{\infty} A_j$ de los conjuntos $A_j = \{x \in \mathbb{R}^2; |x| < 1 - 1/j\}$.

6. Demostrar el teorema 7.1.

7. (a) Probar que de la definición 7.6 de punto de acumulación se deduce que todo entorno de un punto x de acumulación de A contiene, de hecho, infinitos puntos del conjunto A distintos del propio x . **Indicación:** Procédase por reducción al absurdo: sea U un entorno de x y supóngase que $A \cap U$ sólo contiene una cantidad finita de puntos x_1, \dots, x_m distintos de x ; defínase $\delta < \min\{|x_j - x|; j = 1, \dots, m\}$ para llegar a una contradicción con la hipótesis de que x es punto de acumulación.

(b) Demostrar que el supremo y el ínfimo de un conjunto acotado $A \subset \mathbb{R}$ son puntos de adherencia de A . ¿Son siempre de acumulación?

8. Si A y B son dos conjuntos de \mathbb{R}^n tales que $A \subseteq B$, entonces $\text{int } A \subseteq \text{int } B$ y $\bar{A} \subseteq \bar{B}$.

9. Probar que $\text{int } A$ es el mayor conjunto abierto contenido en A , es decir, que si G es un subconjunto abierto de A , entonces $G \subseteq \text{int } A \subseteq A$ y que, en consecuencia, $\text{int } A$ es la unión de todos los subconjuntos abiertos de A . Análogamente, \bar{A} es el menor conjunto cerrado que contiene a A , es decir, si F es un conjunto cerrado tal que $F \supseteq A$, entonces $F \supseteq \bar{A} \supseteq A$, y, de ahí, \bar{A} es la intersección de todos los conjuntos cerrados que contienen a A .

10. Dados dos conjuntos A y B de \mathbb{R}^n probar que

$$\begin{aligned}\overline{A \cup B} &= \overline{A} \cup \overline{B}, \quad \text{int } (A \cap B) = (\text{int } A) \cap (\text{int } B) \\ \overline{A \cap B} &\subseteq \overline{A} \cap \overline{B}, \quad \text{int } (A \cup B) \supseteq (\text{int } A) \cup (\text{int } B)\end{aligned}$$

Encontrar dos conjuntos A y B en \mathbb{R} para los que se tenga

$$\overline{A \cap B} \neq \overline{A} \cap \overline{B}, \quad \text{int } (A \cup B) \neq (\text{int } A) \cup (\text{int } B)$$

11. Probar que si $\overline{A \cap B} = \emptyset$, entonces $\partial(A \cup B) = \partial A \cup \partial B$.

12. Sea $A \subseteq B \subseteq \mathbb{R}^n$. Se dice que A es **denso** en B si $B \subseteq \bar{A}$ (por ejemplo, el conjunto de números racionales \mathbb{Q} es denso en \mathbb{R}). Se pide demostrar las siguientes afirmaciones:

(a) A es denso en B si y sólo si en cada entorno de todo punto de B hay al menos un punto de A , o, lo que es lo mismo, si y sólo si todo punto aislado de B pertenece a A y los demás puntos de B son puntos de acumulación de A .

- (b) Si A es denso en B y B es denso en C , entonces A es denso en C .
- (c) Para cualquier subconjunto abierto A de \mathbb{R}^n se tiene que $\text{int } A \cup \text{ext } A$ es denso en \mathbb{R}^n .

13. Determinar el interior, la frontera, la adherencia y el conjunto de puntos de acumulación de los siguientes conjuntos:

$$(a) \{(x, y); x^2 + y^2 = 1\} \subset \mathbb{R}^2$$

$$(b) \{(x, y); xy < 1\} \subset \mathbb{R}^2$$

$$(c) \{(x, y); 0 \leq x < 1, 0 < y \leq 1\} \subset \mathbb{R}^2$$

$$(d) A = \bigcup_{n \in \mathbb{N}} \{(x, \frac{1}{n}); x \in \mathbb{R}\} \subset \mathbb{R}^2$$

14. Estudiar si son abiertos o cerrados los conjuntos

$$(a) \{(x, y, 0); x^2 + y^2 < 1\} \subset \mathbb{R}^3$$

$$(b) \{(x, y, 0); x^2 + y^2 \leq 1\} \subset \mathbb{R}^3$$

15. Probar que todo conjunto finito (o sea, compuesto de un número finito de elementos) es cerrado.
16. Probar que para cualquier conjunto A de \mathbb{R}^n , el conjunto de sus puntos de acumulación es un conjunto cerrado.
17. La **distancia** entre dos conjuntos A y B de \mathbb{R}^n es el número

$$d(A, B) \stackrel{\text{def}}{=} \inf_{a \in A, b \in B} |a - b|$$

En particular, si $a \in \mathbb{R}^n$ y $B \subset \mathbb{R}^n$, entonces la distancia del punto a al conjunto B es

$$d(a, B) \stackrel{\text{def}}{=} \inf_{b \in B} |a - b|$$

- (i) Probar que si A es compacto y B cerrado, existen sendos puntos $a \in A, b \in B$ tales que $d(A, B) = |a - b|$. Indicación: por la definición de $d(A, B)$, existen sucesiones de puntos $\{a_k\} \subset A, \{b_k\} \subset B$ tales que

$$|a_k - b_k| \rightarrow d(A, B)$$

Encadénense los siguientes hechos:

- $\{a_k\}$ tiene una subsucesión convergente $\{a_{k_m}\}$
- $\{b_{k_m}\}$ es acotada y, por tanto, posee una subsucesión convergente $\{b_{k_{m_j}}\}$
- Los límites $a = \lim a_{k_{m_j}}$ y $b = \lim b_{k_{m_j}}$ pertenecen, respectivamente, a A y B , por ser éstos conjuntos cerrados.

(Nótese que, siendo A compacto y B cerrado, si se tiene $A \cap B = \emptyset$, entonces $d(A, B) > 0$; estúdiase, en relación a este hecho, cuánto vale $d(A, B)$ para $A = \{(0, y); y \geq 0\} \subset \mathbb{R}^2$, $B = \{(x, y); x \geq 0, xy \geq 1\} \subset \mathbb{R}^2$)

- (ii) Sea B un conjunto cerrado de \mathbb{R}^n y sea a un punto de \mathbb{R}^n tal que $a \notin B$; entonces, existe un $b \in B$ tal que $d(a, B) = |a - b| > 0$.
- (iii) Sea A un conjunto abierto y B un conjunto compacto tal que $B \subset A$. Entonces, $d(B, \partial A) > 0$.

18. Probar que son funciones continuas de \mathbb{R}^n en \mathbb{R}

- (i) la norma, $x \mapsto |x|$.
- (ii) La proyección p_j definida por $p_j(x) = x_j$, $j = 1, \dots, n$.

19. Dado un conjunto $A \subset \mathbb{R}^n$, demostrar que la función distancia a A

$$\mathbb{R}^n \ni x \mapsto d(x, A) \stackrel{\text{def}}{=} \inf_{a \in A} |x - a|$$

es uniformemente continua.

20. Sean f, g dos funciones de $D \subseteq \mathbb{R}^n$ en \mathbb{R}^m . Probar las siguientes afirmaciones:

- (i) Si f es continua en $x_0 \in D$ y $f(x_0) \neq 0 \in \mathbb{R}^m$, existe un $\delta > 0$ tal que $f(x) \neq 0$ para todo $x \in B_\delta(x_0) \cap D$.
- (ii) Si f y g son continuas en x_0 y $f(x_0) \neq g(x_0)$, existe un $\alpha > 0$ tal que $f(x) \neq g(x)$ para todo $x \in B_\alpha(x_0) \cap D$.
- (iii) Si f y g son continuas en D y A es el conjunto de puntos $x \in D$ tales que $f(x) \neq g(x)$, entonces existe un conjunto abierto $U \subseteq \mathbb{R}^n$ tal que $A = U \cap D$ (o sea, A es un abierto relativo a D).
- (iv) Si f y g son continuas en D y B es el conjunto de puntos $x \in D$ tales que $f(x) = g(x)$, entonces existe un conjunto cerrado $F \subseteq \mathbb{R}^n$ tal que $B = F \cap D$ (o sea, B es un cerrado relativo a D).
- (v) Si $f(x) = g(x)$ para todo x de un subconjunto denso de D , entonces $f(x) = g(x)$ para todo $x \in D$.

21. Sea $D_0 \subseteq D \subseteq \mathbb{R}^n$ y D_0 denso en D . Si f es una función de D_0 en \mathbb{R}^m , existe a lo más una función $g(x)$ continua en D que sobre D_0 coincida con f . (Se dice que g es una *prolongación* o *extensión* continua de f a D .)
22. Sea $D_0 \subseteq D \subseteq \mathbb{R}^n$ y D_0 denso en D y sea $f : D_0 \rightarrow \mathbb{R}^m$ una función uniformemente continua en D_0 . Probar que entonces existe una función $\bar{f} : D \rightarrow \mathbb{R}^m$ continua en D que coincide con f sobre D_0 ; además, \bar{f} es uniformemente continua. (Indicación: dado $\bar{x} \in D$, sea $\{x_k\}$ una sucesión de puntos de D_0 tal que $x_k \rightarrow \bar{x}$. Por ser f uniformemente continua, se comprueba que conserva el carácter de sucesión de Cauchy de $\{x_k\}$, es decir, que $\{f(x_k)\}$ es una sucesión de Cauchy de puntos de \mathbb{R}^m y, por tanto, convergente a un punto $\bar{y} \in \mathbb{R}^m$. Se define $\bar{f}(\bar{x}) = \bar{y}$ y se comprueba que esta definición no depende de la sucesión elegida para aproximarse a \bar{x} : en efecto, si $\{z_k\}$ es otra sucesión tal que $z_k \rightarrow \bar{x}$, entonces $\{f(z_k)\}$ también es convergente y su límite ha de coincidir con el de $\{f(x_k)\}$ ya que ambas son subsecuencias convergentes de la sucesión que se obtiene al transformar mediante f la sucesión

$$x_1, z_1, x_2, z_2, x_3, z_3, \dots$$

también convergente a \bar{x}).

23. Demostrar el recíproco del resultado establecido en el apartado 7.5, a saber, que si las normas N_1 y N_2 definen la misma topología (o sea, la misma familia de abiertos), entonces existen dos constantes $A > 0$, $B > 0$ tales que

$$N_2(x) \leq N_1(x) \leq B N_2(x)$$

24. Sea $1 < p < \infty$ y sea q tal que $1/p + 1/q = 1$.

(a) Demostrar que

$$|ab| \leq \frac{1}{p} |a|^p + \frac{1}{q} |b|^q$$

para todo par de números $a, b \in \mathbb{R}$. (Indicación: utilizar que la función exponencial $\mathbb{R} \ni x \mapsto e^x$ es convexa y uno a uno de \mathbb{R} sobre $(0, \infty)$).

(b) Sean $x = (x_1, \dots, x_n), y = (y_1, \dots, y_n) \in \mathbb{R}^n$ cualesquiera. Demostrar la desigualdad de Hölder:

$$\sum_{i=1}^n |x_i y_i| \leq \left(\sum_{i=1}^n |x_i|^p \right)^{1/p} \left(\sum_{i=1}^n |y_i|^q \right)^{1/q}$$

y la desigualdad de Minkowski:

$$\left(\sum_{i=1}^n |x_i + y_i|^p \right)^{1/p} \leq \left(\sum_{i=1}^n |x_i|^p \right)^{1/p} + \left(\sum_{i=1}^n |y_i|^p \right)^{1/p}$$

(Indicación: Sea $A = \left(\sum_{i=1}^n |x_i|^p \right)^{1/p}$, $B = \left(\sum_{i=1}^n |y_i|^q \right)^{1/q}$, $X_i = x_i/A$, $Y_i = y_i/B$. Se aplica la parte (a) con $a = X_i$, $b = Y_i$ y se suma en i . Para la desigualdad de Minkowski, se escribe

$$\begin{aligned} \sum_{i=1}^n |x_i + y_i|^p &= \sum_{i=1}^n |x_i + y_i| |x_i + y_i|^{p-1} \leq \\ &\leq \sum_{i=1}^n |x_i| |x_i + y_i|^{p-1} + \sum_{i=1}^n |y_i| |x_i + y_i|^{p-1} \end{aligned}$$

y se aplica la desigualdad de Hölder a cada una de las sumas del último miembro.)

(c) Demostrar que

$$\mathbb{R}^n \ni x \mapsto |x|_p \stackrel{\text{def}}{=} \left(\sum_{i=1}^n |x_i|^p \right)^{1/p}, \quad 1 \leq p < \infty$$

es una norma en \mathbb{R}^n (para $p = 2$ es la norma euclídea y para $p = 1$ es la norma $|\cdot|_1$ que ya conocemos). Esbozar la bola unidad en \mathbb{R}^2 para $p = 3$.

(d) Demostrar que, para cada $x \in \mathbb{R}^n$ fijado, $|x|_p \geq |x|_{p'}$ si $p' > p$ y que

$$\lim_{p \rightarrow \infty} |x|_p = |x|_\infty$$

(y de ahí la notación $|x|_\infty$)

(e) Para cada base \mathfrak{B} de \mathbb{R}^n estará definida la familia de normas $|\cdot|_p$ utilizando las coordenadas respecto a \mathfrak{B} en lugar de las coordenadas canónicas (x_1, \dots, x_n) . Calcular las normas $|x|_p$, $1 \leq p \leq \infty$ del vector $(2, 2) \in \mathbb{R}^2$ relativas a la base canónica y a la base formada por los vectores $(1, 0)$ y $(1, 2)$.

25. Demostrar que todo subespacio vectorial E de \mathbb{R}^n es un conjunto cerrado (Indicación: Identificar E con un espacio \mathbb{R}^p apropiado y aplicar el teorema de Bolzano-Weierstrass.).

26. En el apartado 1.4 definimos como *producto interior* o *escalar* en \mathbb{R}^n toda aplicación $\mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$ que es *simétrica*, *bilineal* y *definida positiva*; se denota por $\langle x, y \rangle \mapsto \langle x, y \rangle$.

- (a) Dado un producto interior, demostrar que $x \mapsto \langle x, x \rangle^{1/2}$ es una norma (**Indicación:** la demostración es la misma que para el producto escalar ordinario; véase el apartado 1.4.)
- (b) Si $\langle x, y \rangle$ es un producto interior en \mathbb{R}^n y $|x| = \langle x, x \rangle^{1/2}$ la norma asociada a él, demostrar que para todo $x, y \in \mathbb{R}^n$ se satisface la "ley del paralelogramo"

$$|x+y|^2 + |x-y|^2 = 2(|x|^2 + |y|^2)$$

("la suma de los cuadrados de las diagonales de un paralelogramo es igual a la suma de los cuadrados de sus lados").

- (c) Recíprocamente, demostrar que una norma en \mathbb{R}^n proviene de un producto escalar si se satisface dicha ley del paralelogramo. (**Indicación:** Definase

$$\langle x, y \rangle = \frac{1}{4} (|x+y|^2 - |x-y|^2)$$

La simetría es inmediata. Para probar la linealidad $\langle x+y, z \rangle = \langle x, z \rangle + \langle y, z \rangle$, desarróllense ambos miembros y utilícese la ley del paralelogramo. Para probar que $\langle cx, y \rangle = c \langle x, y \rangle$ se comienza con c racional, observando que, por lo anterior (procediendo por inducción), si $m, n \in \mathbb{N}$

$$m \langle x, y \rangle = \langle mx, y \rangle = \left\langle n \frac{m}{n} x, y \right\rangle = n \left\langle \frac{m}{n} x, y \right\rangle$$

También se tiene $\langle -x, y \rangle = -\langle x, y \rangle$ por las propiedades de la norma, con lo que $\langle cx, y \rangle = c \langle x, y \rangle$ para todo $c \in \mathbb{Q}$. Se apela ahora a la continuidad de las funciones

$$\mathbb{R} \ni \alpha \mapsto \alpha x \in \mathbb{R}^n \text{ para } x \in \mathbb{R}^n \text{ fijo}$$

y

$$\mathbb{R}^n \ni x \mapsto \langle x, y \rangle \text{ para } y \in \mathbb{R}^n \text{ fijo}$$

que deriva de la continuidad de la norma, para concluir que $\langle cx, y \rangle = c \langle x, y \rangle$ para todo $c \in \mathbb{R}$.

27. Probar que si N_1 y N_2 son dos normas en \mathbb{R}^n , entonces

$$N(x) = tN_1(x) + (1-t)N_2(x)$$

también es norma en \mathbb{R}^n .

28. Estudiar cuáles de las siguientes fórmulas definen una norma en \mathbb{R}^2 :

$$(a) (x^2 + xy + y^2)^{1/2} \quad (b) (x^2 - 3xy + y^2)^{1/2}$$

$$(c) (|x| + |y|)^{1/2} \quad (d) \frac{1}{4}(|x| + |y|) + \frac{3}{4}(x^2 + y^2)^{1/2}$$

29. Sea N una norma cualquiera en $\mathcal{L}(\mathbb{R}^n)$. Probar que existe una constante k tal que $N(ST) \leq kN(S)N(T)$ para todo par de operadores lineales S, T .

30. La norma uniforme de una aplicación lineal $L \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ depende de la norma considerada en \mathbb{R}^n y \mathbb{R}^m .

- (a) Se considera en \mathbb{R}^n y \mathbb{R}^m la norma $|\cdot|_\infty$. Entonces

$$\|L\| = \max_{1 \leq i \leq m} \sum_{j=1}^n |a_{ij}|$$

siendo $A = (a_{ij})$ la matriz representativa de L respecto a las bases de \mathbb{R}^n y \mathbb{R}^m consideradas (normalmente, las respectivas bases canónicas), es decir, $\|L\|$ es la mayor de las normas $|\cdot|_1$ de los vectores fila de la matriz.

- (b) Se considera en \mathbb{R}^n y \mathbb{R}^m la norma $|\cdot|_1$. Entonces

$$\|L\| = \max_{1 \leq j \leq n} \sum_{i=1}^m |a_{ij}|$$

es decir, $\|L\|$ es la mayor de las normas $|\cdot|_1$ de los vectores columna de la matriz.

(Indicación: Probar primero, en ambos casos, que $\|L\|$ es menor o igual que el número indicado y después que se alcanza la igualdad $\|L\| = |Lx|$ en un cierto punto de $B_1(0)$.)

31. Se considera en \mathbb{R}^n y \mathbb{R}^m la norma euclídea. Entonces

- (a) Si $L \in \mathcal{L}(\mathbb{R}^n)$ y la matriz A , representante de L en la base canónica de \mathbb{R}^n , es simétrica, se tiene

$$\|L\| = |\lambda_1|$$

siendo λ_1 el autovalor de A de mayor valor absoluto. Probar que se tiene

$$\|L\| = \max_{|x|=1} |\langle Ax, x \rangle| = \max_{|x|=1} \left| \sum_{i=1}^n \left(\sum_{j=1}^n a_{ij} x_j \right) x_i \right|$$

- (b) En el caso general de $L \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$,

$$\|L\| = \sqrt{\Lambda_1}$$

siendo Λ_1 el mayor autovalor de la matriz $A^T A$.

(Indicación: (a) Como A es simétrica, sus autovalores son reales y, como vimos al final del apartado 6.2,

$$\max_{|x|=1} |\langle Ax, x \rangle| = |\lambda_1|$$

Los autovalores de A^2 son los cuadrados de los autovalores de A . Entonces

$$\begin{aligned} \|L\|^2 &= \max_{|x|=1} |Lx|^2 = \max_{|x|=1} \langle Ax, Ax \rangle = \max_{|x|=1} \langle A(Ax), x \rangle = \\ &= \max_{|x|=1} \langle A^2 x, x \rangle = \lambda_1^2 \end{aligned}$$

Para (b), recuérdese que $\langle Ax, Ax \rangle = \langle A^T A x, x \rangle$ y que $A^T A$ es una matriz simétrica positiva.)

32. **Exponencial de una matriz.** Demostrar que si A es una matriz $n \times n$ la serie

$$\sum_{k=0}^{\infty} \frac{A^k}{k!}$$

es (absolutamente) convergente en $\mathcal{L}(\mathbb{R}^n) \approx \mathbb{R}^{n^2}$. Su suma se representa por e^A .

(Indicación: Utilícese (7.44) y el criterio de comparación de Weierstrass.)

Supongamos que $A \in \mathcal{L}(\mathbb{R}^n)$ deja invariante un subespacio $E \subset \mathbb{R}^n$ (es decir, $Ax \in E$ para todo $x \in E$). Probar que e^A también deja invariante E .

33. (a) Probar que $\|T\| \cdot \|T^{-1}\| \geq 1$ para todo $T \in \mathcal{L}(\mathbb{R}^n)$ invertible.
 (b) Si T tiene dos autovalores reales λ, μ tales que $|\lambda| \neq |\mu|$, entonces $\|T\| \cdot \|T^{-1}\| > 1$.

34. Demostrar que si $T \in \mathcal{L}(\mathbb{R}^n)$ es tal que $\|T - I\| < 1$, entonces T es invertible y la serie $\sum_{k=0}^{\infty} (I - T)^k$ converge absolutamente a T^{-1} . Hallar una cota superior para $\|T^{-1}\|$. (I es el operador identidad en \mathbb{R}^n .)

(Indicación: la idea consiste en inspirarse en la serie numérica convergente

$$\frac{1}{1-x} = \sum_{k=0}^{\infty} x^k \text{ para } 0 < x < 1 \text{ con } x = \|I - T\|$$

para probar que la serie indicada converge absolutamente a un $S \in \mathcal{L}(\mathbb{R}^n)$. Después hay que probar que $TSx = x$ para todo $x \in \mathbb{R}^n$.)

35. Sea $T_0 \in \mathcal{L}(\mathbb{R}^n)$ invertible. Hallar $\varepsilon > 0$ tal que si $\|T - T_0\| < \varepsilon$, entonces T es invertible. (El conjunto de operadores invertibles en $\mathcal{L}(\mathbb{R}^n)$ es abierto en $\mathcal{L}(\mathbb{R}^n) \approx \mathbb{R}^{n^2}$; en otras palabras, si una matriz A tiene inversa, las matrices con norma suficientemente próxima a la de A también tienen inversa.)

(Indicación: Intentar aplicar a $T_0^{-1}T$ el resultado del problema anterior.)

36. Sea $f: S \rightarrow \mathbb{R}$ continua en un conjunto conexo $S \subseteq \mathbb{R}^n$. Probar que si para un cierto número real a se tiene que $f(x) \neq a$ para todo $x \in S$, entonces se verifica que o bien $f(x) > a$ para todo $x \in S$ o bien $f(x) < a$ para todo $x \in S$.

37. Demostrar el siguiente *teorema de punto fijo*: Sea $I = [0, 1]$ y sea $f: I \rightarrow \mathbb{R}$ una función continua tal que $f(I) \subseteq I$; entonces, existe $p \in I$ tal que $f(p) = p$.

(Indicación: (i) Hágase en primer lugar una figura que muestre el significado geométrico del teorema: la gráfica de f corta a la diagonal Δ del cuadrado unidad $I \times I = \{(x, y) \in \mathbb{R}^2; 0 \leq x \leq 1, 0 \leq y \leq 1\}$.

(ii) Se puede suponer $f(0) > 0$ y $f(1) < 1$ (¿por qué?); defínase $F: I \rightarrow \mathbb{R}^2$ por $F(x) = (x, f(x))$; compruébese que también es continua y que, por tanto, $F(I)$ (la gráfica de f) es conexo; si no se cumple la tesis del teorema, los conjuntos $A_1 = \{(x, y) \in \mathbb{R}^2; x < y\}$, $A_2 = \{(x, y) \in \mathbb{R}^2; x > y\}$ producen una "desconexión" de $F(I)$.)

38. Sea $A \subset \mathbb{R}^n$ y supongamos que hay un arco (continuo) $\varphi: [a, b] \rightarrow \mathbb{R}^n$ que une $p \in \text{int}(A)$ con $q \in \text{ext}(A)$. Probar que existe $t^* \in (a, b)$ tal que $\varphi(t^*) \in \partial A$: "para salir de un conjunto hay que atravesar su frontera". (Indicación: Proceder por contradicción como en el problema anterior utilizando la conexión del arco.)
39. Sea \mathcal{F} una familia de conjuntos conexos tal que $\bigcap \{S; S \in \mathcal{F}\} \neq \emptyset$. Entonces, $\bigcup \{S; S \in \mathcal{F}\}$ también es conexo. Aplicar este resultado para probar que \mathbb{R}^n es conexo utilizando el hecho de que las rectas en \mathbb{R}^n son conjuntos conexos. (Indicación:
- (i) Si $A = \bigcup \{S; S \in \mathcal{F}\}$ no es conexo, existen G_1, G_2 abiertos de \mathbb{R}^n tales que $G_1 \cap A, G_2 \cap A$ son conjuntos no vacíos disjuntos cuya unión es A .
 - (ii) Sea $x \in \bigcap \{S; S \in \mathcal{F}\}$ y supongamos que, por ejemplo, $x \in G_1 \cap A$; por la hipótesis (i), hay al menos un $S_i \in \mathcal{F}$ tal que $G_2 \cap S_i \neq \emptyset$. También se tiene $G_1 \cap S_i \neq \emptyset$ (pues contiene a x) y, por (i), $(G_1 \cap S_i) \cup (G_2 \cap S_i) = S_i$. Se llega a una contradicción con la conexión de S_i .
 - (iii) \mathbb{R}^n es la unión de las rectas que pasan por el origen.)
40. Probar que si $A \subset \mathbb{R}^n$ es conexo y B es un conjunto tal que $A \subset B \subseteq \bar{A}$, entonces B es conexo. En particular, la adherencia de un conjunto conexo es conexa. (Indicación: Se razona, de nuevo, por contradicción:
- (i) Si B no es conexo, existen G_1, G_2 abiertos de \mathbb{R}^n tales que $G_1 \cap B, G_2 \cap B$ son conjuntos no vacíos disjuntos cuya unión es B .
 - (ii) Como A es conexo y está contenido en B , o bien $A \cap G_1 = \emptyset$ o bien $A \cap G_2 = \emptyset$, y, por tanto, o bien $A \subset G_2$ o bien $A \subset G_1$.
 - (iii) Supongamos que $A \cap G_2 = \emptyset$; entonces, también se verifica $\bar{A} \cap G_2 = \emptyset$, pues si existe un punto de acumulación p de A tal que $p \in G_2$, se tendría, por ser G_2 abierto, $A \cap G_2 \neq \emptyset$ y se llega a una contradicción.
 - (iv) Pero si $\bar{A} \cap G_2 = \emptyset$, también será $B \cap G_2 = \emptyset$ contra la hipótesis hecha en (i).)
41. Nos proponemos mostrar que si un conjunto A no es conexo, se puede expresar como unión de partes conexas. Una **componente conexa** de A es un subconjunto conexo maximal de A . Es decir, C es una componente conexa de A si C es conexo y no existe un subconjunto conexo de A que contenga propiamente a C . Si A es conexo, su única componente

conexa es, naturalmente, él mismo. Si todas las componentes conexas de A se reducen a un único punto, se dice que A es *totalmente discontinuo*.

- (i) Cada punto $x \in A$ está contenido en una única componente conexa de A (que se llama componente conexa de x en A), y, en consecuencia, A es la unión disjunta de sus componentes. (Indicación: para cada $x \in A$, considérese la unión de todos los subconjuntos conexos de A que contienen a x y utilícese el problema 39.)
 - (ii) Las componentes conexas de A son conjuntos cerrados relativos a A . (Indicación: problema 40.)
 - (iii) Si A es abierto, las componentes conexas de A son conjuntos abiertos relativos a A . (Indicación: las bolas $B_\varepsilon(x)$ son conjuntos conexos.)
42. Sean A y B dos conjuntos conexos de \mathbb{R}^n tales que $\bar{A} \cap B \neq \emptyset$ (se dice en este caso que A y B no están separados). Probar que $A \cup B$ es conexo. En particular, si $A \cap B \neq \emptyset$, $A \cup B$ es conexo. (Indicación: Imitar los razonamientos del problema 40 para llegar a que

$$\begin{aligned} G_1 \cap (A \cup B) &= (G_1 \cap A) \cup (G_1 \cap B) = \emptyset \\ &\quad \text{o bien} \\ G_2 \cap (A \cup B) &= (G_2 \cap A) \cup (G_2 \cap B) = \emptyset \end{aligned}$$

en contradicción con la hipótesis de que $A \cup B$ no sea conexo.)

43. Estudiar la conexión de los siguientes conjuntos:

- (i) $\{(x, y); x^2 - y^2 \geq 1\}$
- (ii) $\{x \in \mathbb{R}^n; 0 < |x| \leq 1\}$
- (iii) $\{x \in \mathbb{R}^n; |x| = 1\}$
- (iv) $\{x \in \mathbb{R}^n; 1 < |x| \leq 2\}$
- (v) $\{(x, y); \max(|x|, |y|) < 1\} \cup \{(x, y); (x-2)^2 + y^2 < 1\}$
- (vi) $\{(x, y); \max(|x|, |y|) \leq 1\} \cup \{(x, y); (x-2)^2 + y^2 < 1\}$

44. Se considera el conjunto $\{0, 1\}$ con la topología inducida por la usual de \mathbb{R} (nótese que se trata de la topología discreta: todas las partes son abiertas). Sea $S \subset \mathbb{R}^n$. Demostrar que S es conexo si y sólo si toda función $f: S \rightarrow \{0, 1\}$ continua es constante (naturalmente, se podrían tomar dos puntos cualesquiera distintos de \mathbb{R} en lugar de 0 y 1).

Aplicar lo anterior para hacer demostraciones alternativas (y más simples que las allí sugeridas) de los resultados de los problemas 39, 40 y 42.

Capítulo 8

Funciones diferenciables

Este capítulo trata fundamentalmente de la generalización a funciones vectoriales de la noción clave de *diferenciabilidad*, es decir, de la propiedad de aproximación local de una función no lineal por una lineal.

Extenderemos, pues, en los apartados 8.1 y 8.2, definiciones y resultados que vimos con detalle en los capítulos 3 y 4 para funciones escalares, aplicándolos en buena parte de los casos a las *componentes* f_i de una función vectorial $f = (f_1, \dots, f_m)$ para obtener el correspondiente enunciado para f .

En el apartado 8.3 se tratan algunos aspectos *dinámicos* de la teoría de curvas basados en el cálculo diferencial, sin entrar en otros aspectos más geométricos, como la longitud de un arco, que se basan en el cálculo integral. En 8.4 se exponen algunas cuestiones que abren el camino al cálculo diferencial en espacios abstractos.

8.1 Diferencial de una función

Definición 8.1 Sea $f : D \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^m$ y sea x_0 un punto interior de D . Se dice que f es **diferenciable** en x_0 si existe una aplicación lineal $L : \mathbb{R}^n \rightarrow \mathbb{R}^m$, dependiente, en general, de x_0 tal que

$$\lim_{h \rightarrow 0} \frac{1}{|h|} [f(x_0 + h) - f(x_0) - L(h)] = 0 \quad (8.1)$$

Para $m = 1$, se tiene, naturalmente, la definición del capítulo 3. Se dice que f es **diferenciable en un conjunto abierto** D si lo es en cada punto de D y que es **diferenciable en un conjunto** $A \subset \mathbb{R}^n$ si es diferenciable en un conjunto abierto D que contiene a A .

La propiedad de diferenciabilidad se define a través de un límite y, por tanto, depende únicamente de las topologías de \mathbb{R}^n y \mathbb{R}^m , las cuales, como

sabemos, son las mismas cualquiera que sea la norma considerada en dichos espacios. La definición 8.1, en consecuencia, es independiente de la norma que deseemos utilizar en uno y otro espacio. Además, hay a lo más una aplicación lineal que satisface (8.1). En efecto, si $L_1, L_2 \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ satisfacen ambas (8.1), entonces

$$\lim_{h \rightarrow 0} \frac{1}{|h|} [L_1(h) - L_2(h)] = 0$$

de donde, tomando un $x \in \mathbb{R}^n$, $x \neq 0$, fijo, se deduce

$$0 = \lim_{t \rightarrow 0} \frac{|L_1(tx) - L_2(tx)|}{|tx|} = \frac{|L_1(x) - L_2(x)|}{|x|}$$

y, en consecuencia, $L_1(x) = L_2(x)$ para todo $x \in \mathbb{R}^n$, es decir, $L_1 = L_2$.

Si f es diferenciable en x_0 , la única aplicación lineal L que satisface (8.1) se llama *diferencial de f en x_0* y se denota por $Df(x_0)$ o $f'(x_0)$. Se expresa también (8.1) escribiendo

$$f(x_0 + h) - f(x_0) = Df(x_0) \cdot h + g(x_0; h) \quad (8.2)$$

con $g \stackrel{\text{def}}{=} f(x_0 + h) - f(x_0) - Df(x_0) \cdot h$ tal que

$$\lim_{h \rightarrow 0} \frac{|g(x_0; h)|}{|h|} = 0 \quad (8.3)$$

o, equivalentemente (poniendo $x = x_0 + h$)

$$f(x) - f(x_0) = Df(x_0) \cdot (x - x_0) + g(x_0; x) \quad (8.4)$$

con

$$\lim_{x \rightarrow x_0} \frac{|g(x_0; x)|}{|x - x_0|} = 0 \quad (8.5)$$

lo que se expresa abreviadamente escribiendo

$$|g(x_0; x)| = o(|x - x_0|) \text{ cuando } x \rightarrow x_0 \quad (8.6)$$

Si $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ es lineal, $Df(x_0) = f$ para todo $x_0 \in \mathbb{R}^n$, pues, obviamente,

$$\lim_{h \rightarrow 0} \frac{f(x_0 + h) - f(x_0) - f(h)}{|h|} = \lim_{h \rightarrow 0} \frac{f(x_0) + f(h) - f(x_0) - f(h)}{|h|} = 0$$

De la definición 8.1 y de las propiedades elementales de los límites se deduce inmediatamente que si $f: D \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^m$ y $\varphi: D \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^m$ son

diferenciables en $x_0 \in D$, entonces $f + \varphi$ y αf ($\alpha \in \mathbb{R}$) también lo son y se tiene

$$\begin{aligned} D(f + \varphi)(x_0) &= Df(x_0) + D\varphi(x_0) \\ D(\alpha f)(x_0) &= \alpha(Df(x_0)) \end{aligned} \quad (8.7)$$

Es fácil identificar la matriz $m \times n$ asociada a la diferencial $Df(x_0)$:

Proposición 8.1 (i) La función $f = (f_1, \dots, f_m) : D \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^m$ es diferenciable en $x_0 \in D$ si y sólo si lo es cada una de sus componentes.

(ii) Si f es diferenciable en x_0 , $Df(x_0)$ está dada por la matriz de derivadas parciales

$$\left(\frac{\partial f_i}{\partial x_j}(x_0) \right), \quad i = 1, \dots, m; \quad j = 1, \dots, n \quad (8.8)$$

(la matriz jacobiana de (f_1, \dots, f_m) respecto a (x_1, \dots, x_n) evaluada en $(x_1, \dots, x_n) = (x_1^0, \dots, x_n^0)$).

Demostración: los argumentos clave son la proposición 7.6 del capítulo anterior y la expresión de la diferencial de una función escalar que conocemos del capítulo 3. Si f es diferenciable en x_0 y $Df(x_0)$ está dada por la matriz $(a_{ij}(x_0))$, $i = 1, \dots, m$, $j = 1, \dots, n$, entonces, desplegando (8.2) se tiene

$$\begin{aligned} f_1(x_0 + h) - f_1(x_0) &= a_{11}(x_0)h_1 + \dots + a_{1n}(x_0)h_n + g_1(x_0; h) \\ &\vdots \\ f_m(x_0 + h) - f_m(x_0) &= a_{m1}(x_0)h_1 + \dots + a_{mn}(x_0)h_n + g_m(x_0; h) \end{aligned} \quad (8.9)$$

y de (8.3), por la proposición 7.6,

$$\lim_{h \rightarrow 0} \frac{g_i(x_0; h)}{|h|} = 0, \quad i = 1, \dots, m \quad (8.10)$$

Esto implica que las componentes f_i son diferenciables y, como sabemos del capítulo 3, que

$$a_{ij}(x_0) = \frac{\partial f_i}{\partial x_j}(x_0) \quad (8.11)$$

Recíprocamente, si las componentes f_i son diferenciables entonces se tiene

$$\begin{aligned} f_1(x_0 + h) - f_1(x_0) &= \frac{\partial f_1}{\partial x_1}(x_0)h_1 + \dots + \frac{\partial f_1}{\partial x_n}(x_0)h_n + g_1(x_0; h) \\ &\vdots \\ f_m(x_0 + h) - f_m(x_0) &= \frac{\partial f_m}{\partial x_1}(x_0)h_1 + \dots + \frac{\partial f_m}{\partial x_n}(x_0)h_n + g_m(x_0; h) \end{aligned} \quad (8.12)$$

con g_1, \dots, g_m verificando (8.10), lo que, de nuevo por la proposición 7.6, implica que $g = (g_1, \dots, g_m)$ verifica (8.3) y que, en consecuencia, la aplicación lineal cuya matriz es

$$\left(\frac{\partial f_i}{\partial x_j}(x_0) \right), \quad i = 1, \dots, m; \quad j = 1, \dots, n$$

verifica (8.1), es decir, que f es diferenciable en x_0 y $Df(x_0)$ está dada por la matriz jacobiana (8.8).

(La equivalencia " f diferenciable $\Leftrightarrow f_i$ diferenciables" permitiría definir como función diferenciable aquella cuyas componentes son diferenciables, y se tendría la unicidad de la diferencial a partir de la unicidad de la diferencial para las funciones escalares que vimos en el capítulo 3. La definición 8.1 y la demostración de la unicidad que hemos dado tienen la ventaja de que son válidas en situaciones más generales como la que consideramos en la nota 2 de final de capítulo.)

Si $m = 1$, o sea, si se trata de una función escalar, la matriz jacobiana es la matriz fila

$$\left(\frac{\partial f}{\partial x_1}(x_0), \dots, \frac{\partial f}{\partial x_n}(x_0) \right) = \nabla f(x_0)$$

Si $n = 1$, las componentes f_i son funciones de una variable y la matriz jacobiana es la matriz columna de derivadas ordinarias

$$\begin{pmatrix} \frac{df_1}{dx} \\ \vdots \\ \frac{df_m}{dx} \end{pmatrix}$$

Proposición 8.2 Si la función $f : D \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^m$ es diferenciable en $x_0 \in D$ entonces es continua en x_0 .

En efecto, basta tener en cuenta el correspondiente resultado para las componentes de f y la proposición precedente. Más precisamente, se tiene:

Proposición 8.3 Sea $f : D \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^m$ diferenciable en $x_0 \in D$. Entonces, para todo $\varepsilon > 0$ existe $\delta > 0$ tal que $B_\delta(x_0) \subseteq D$ y

$$|f(x) - f(x_0)| \leq (\|Df(x_0)\| + \varepsilon) |x - x_0| \quad (8.13)$$

para todo $x \in B_\delta(x_0)$.

Demostración: Como x_0 se supone un punto interior de D , existe $\delta_1 > 0$ tal que $B_{\delta_1}(x_0) \subseteq D$. Tomando normas en (8.4):

$$\begin{aligned} |f(x) - f(x_0)| &\leq |Df(x_0) \cdot (x - x_0)| + |g(x_0; x)| \leq \\ &\leq \|Df(x_0)\| |x - x_0| + |g(x_0; x)| \end{aligned} \quad (8.14)$$

Pero, por (8.5), dado $\varepsilon > 0$, existe $\delta > 0$, $\delta \leq \delta_1$, tal que, si $|x - x_0| < \delta$, entonces

$$|g(x_0; x)| \leq \varepsilon |x - x_0| \quad (8.15)$$

lo que, sustituido en (8.14), da (8.13). (De (8.13) se deduce inmediatamente la continuidad de f en x_0 . Nótese: para obtener la desigualdad (8.13) se supone que se han fijado unas normas en \mathbb{R}^n y \mathbb{R}^m , las que se deseen, y que $\|\cdot\|$ representa, de acuerdo con el convenio adoptado en el apartado 7.6, la norma uniforme en $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ inducida por ellas.)

La condición suficiente de diferenciabilidad que establecimos en el teorema 3.2 se traslada a través de sus componentes, gracias a la proposición 8.1, a una función vectorial:

Teorema 8.1 Sea $D \subset \mathbb{R}^n$ un conjunto abierto y sea la función $f = (f_1, \dots, f_m) : D \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^m$. Si las derivadas parciales $\partial f_i / \partial x_j$ existen y son continuas en D , entonces f es diferenciable en D .

(Si $n = 1$, como las componentes f_i son funciones de una variable, basta la existencia de las derivadas ordinarias df_i/dx para garantizar que aquellas son funciones diferenciables y que lo mismo ocurrirá con f .)

El teorema 8.1 motiva la siguiente

Definición 8.2 Sea $D \subseteq \mathbb{R}^n$ abierto y sea $f = (f_1, \dots, f_m) : D \rightarrow \mathbb{R}^m$. Se dice que f es de clase uno (o de clase C^1) en D si lo son las componentes f_1, \dots, f_m (definición 3.9), es decir, si existen las derivadas parciales $\partial f_i / \partial x_j$ y son funciones continuas en todo D . Se denota por $C^1(D)$ al conjunto de tales funciones. Si A es un subconjunto de \mathbb{R}^n , se dice que f es de clase uno (o de clase C^1) en A (y se escribe $f \in C^1(A)$) si es de clase uno en un conjunto abierto D que contiene a A .

Pongamos por otra parte:

Definición 8.3 Se dice que $f : D \rightarrow \mathbb{R}^m$ es diferenciable con continuidad o continuamente diferenciable en D si:

1. f es diferenciable en D .

2. La aplicación $\mathbb{R}^n \supseteq D \ni x_0 \mapsto Df(x_0) \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ es continua.

En realidad, las nociones " $f \in C^1(D)$ " y " f continuamente diferenciable en D " son equivalentes. En efecto, si $f \in C^1(D)$, entonces, por el teorema 8.1, f es diferenciable en D y, además, utilizando, por ejemplo, en $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ la norma

$$|A|_\infty = \max \{|a_{ij}| : i = 1, \dots, m, j = 1, \dots, n\}$$

(para las matrices $A = (a_{ij})$ representativas) se tiene que

$$|Df(a) - Df(b)|_\infty = \max_{i,j} \left| \frac{\partial f_i}{\partial x_j}(a) - \frac{\partial f_i}{\partial x_j}(b) \right| \quad (8.16)$$

y, en consecuencia, por la continuidad de las derivadas parciales, que la aplicación $x_0 \mapsto Df(x_0)$ es continua. Recíprocamente, si f es continuamente diferenciable, f es, en primer lugar, diferenciable, y, por tanto, existen las derivadas parciales $\partial f_i / \partial x_j$, y éstas son continuas por (8.16) y la continuidad de $x_0 \mapsto Df(x_0)$.

8.2 La regla de la cadena y sus consecuencias

La regla de la cadena del cálculo diferencial establece que "la diferencial de una composición de funciones

$$\mathbb{R}^n \xrightarrow{f} \mathbb{R}^m \xrightarrow{g} \mathbb{R}^p$$

es la composición $Dg \circ Df$ de las diferenciales". Con precisión: sean $D \subseteq \mathbb{R}^n$ y $\Omega \subseteq \mathbb{R}^m$ dos conjuntos abiertos y sean

$$f : D \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^m \text{ y } g : \Omega \subseteq \mathbb{R}^m \rightarrow \mathbb{R}^p$$

dos funciones continuas. Supongamos que $x_0 \in D$ es tal que $f(x_0) \in \Omega$; entonces, $D' = f^{-1}(\Omega) \subseteq D$ es un conjunto abierto de \mathbb{R}^n que contiene a x_0 y en D' está definida la aplicación compuesta $g \circ f : D' \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^p$ por $(g \circ f)(x) = g(f(x))$. Pues bien, bajo estas hipótesis se tiene:

Teorema 8.2 Si f es diferenciable en $x_0 \in D$ y g es diferenciable en $y_0 = f(x_0)$, entonces $h = g \circ f$ es diferenciable en x_0 y se tiene

$$Dh(x_0) = Dg(y_0) \circ Df(x_0) \quad (8.17)$$

Demostración: Por hipótesis

$$f(x) - f(x_0) = Df(x_0) \cdot (x - x_0) + r_1(x_0; x) \quad (8.18)$$

$$g(y) - g(y_0) = Dg(y_0) \cdot (y - y_0) + r_2(y_0; y) \quad (8.19)$$

con $|r_1(x_0; x)| = o(|x - x_0|)$ y $|r_2(y_0; y)| = o(|y - y_0|)$. Se trata de probar que

$$h(x) - h(x_0) - (Dg(f(x_0)) \circ Df(x_0)) \cdot (x - x_0) = o(|x - x_0|) \quad (8.20)$$

Se tiene:

$$\begin{aligned} h(x) - h(x_0) &= g(f(x)) - g(f(x_0)) = (\text{por (8.19)}) = \\ &= Dg(f(x_0)) \cdot (f(x) - f(x_0)) + r_2(f(x_0); f(x)) = (\text{por (8.18)}) = \\ &= Dg(f(x_0)) \cdot [Df(x_0) \cdot (x - x_0) + r_1(x_0; x)] + r_2(f(x_0); f(x)) = \\ &= (Dg(f(x_0)) \text{ es lineal}) = \\ &= (Dg(f(x_0)) \circ Df(x_0)) \cdot (x - x_0) + Dg(f(x_0)) \cdot r_1(x_0; x) + r_2(f(x_0); f(x)) \end{aligned}$$

Bastará comprobar entonces que el segundo y tercer sumandos del último término son $o(|x - x_0|)$. Para el segundo se tiene:

$$\frac{|Dg(f(x_0)) \cdot r_1(x_0; x)|}{|x - x_0|} \leq \|Dg(f(x_0))\| \frac{|r_1(x_0; x)|}{|x - x_0|} \xrightarrow{x \rightarrow x_0} 0$$

y en cuanto al tercero, notemos, en primer lugar, que, por (8.13), existe $\delta_1 > 0$ tal que, si $|x - x_0| < \delta_1$,

$$|f(x) - f(x_0)| \leq M|x - x_0|, \quad M = \|Dg(f(x_0))\| + 1 \quad (8.21)$$

Dado $\varepsilon > 0$, existe $\eta > 0$ tal que, si $|y - y_0| < \eta$,

$$|r_2(y_0; y)| < \frac{\varepsilon}{M}|y - y_0| \quad (8.22)$$

Entonces, si x es tal que $|x - x_0| < \delta = \min(\delta_1, \eta/M)$ se tiene, por (8.21),

$$|f(x) - f(x_0)| \leq M \frac{\eta}{M} = \eta$$

y, en consecuencia, por (8.22) y (8.21),

$$|r_2(f(x_0); f(x))| < \frac{\varepsilon}{M}|f(x) - f(x_0)| \leq \frac{\varepsilon}{M}M|x - x_0| = \varepsilon|x - x_0|$$

es decir, $|r_2(f(x_0); f(x))| = o(|x - x_0|)$ como queríamos demostrar.

La diferencial $Dh(x_0) = Dg(y_0) \circ Df(x_0)$ estará representada por el producto de matrices

$$\left(\frac{\partial h_k}{\partial x_j}(x_0) \right) = \left(\frac{\partial g_k}{\partial y_i}(f(x_0)) \right) \left(\frac{\partial f_i}{\partial x_j}(x_0) \right) \quad \begin{matrix} i = 1, \dots, m \\ j = 1, \dots, n \\ k = 1, \dots, p \end{matrix} \quad (8.23)$$

que es una matriz $p \times n$.

Si, en particular, $p = 1$, entonces g y h son escalares y, simplificando la escritura, de (8.23) resulta

$$\left(\frac{\partial h}{\partial x_1}, \dots, \frac{\partial h}{\partial x_n} \right) = \left(\frac{\partial g}{\partial y_1}, \dots, \frac{\partial g}{\partial y_m} \right) \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \dots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \dots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \dots & \frac{\partial f_m}{\partial x_n} \end{pmatrix}$$

o sea,

$$\frac{\partial h}{\partial x_j} = \frac{\partial g}{\partial y_1} \frac{\partial f_1}{\partial x_j} + \dots + \frac{\partial g}{\partial y_m} \frac{\partial f_m}{\partial x_j}, \quad j = 1, \dots, n \quad (8.24)$$

que es la regla de la cadena para derivadas parciales (apartado 4.1).

Si $n = 1$, la regla de la cadena toma la forma

$$(g \circ f)'(x_0) = Dg(f(x_0)) \cdot f'(x_0) \quad (8.25)$$

$(Dg(f(x_0))) \in \mathcal{L}(\mathbb{R}^m, \mathbb{R}^p)$ "es" una matriz $p \times m$, $f'(x_0)$ un vector de \mathbb{R}^m y $(g \circ f)'(x_0)$ queda identificada a un vector de \mathbb{R}^p .

Si también es $p = 1$, entonces $(g \circ f)'(x_0)$ se identifica con un número real que vale

$$(g \circ f)'(x_0) = \langle \nabla g(f(x_0)), f'(x_0) \rangle \quad (8.26)$$

Supongamos que $p = 1$ y que $f = (f_1, \dots, f_m)$ y g son de clase C^1 en D y Ω , respectivamente. Entonces son diferenciables y la fórmula (8.24) se aplica en todo punto $x_0 \in D$ y el correspondiente $y_0 = f(x_0) \in \Omega$:

$$\begin{aligned} (D_j h)(x_0) &= \\ &= ((D_1 g) \circ f)(x_0) \cdot (D_j f_1)(x_0) + \dots + ((D_m g) \circ f)(x_0) \cdot (D_j f_m)(x_0), \\ &\qquad\qquad\qquad j = 1, \dots, n \end{aligned} \quad (8.27)$$

fórmula en la que se aprecia que $D_j h$ es el resultado de sumas, productos y composiciones de funciones continuas, por lo que será ella misma continua y, en consecuencia, $h = g \circ f$ es de clase C^1 en D' . En el caso general, si se supone que $f = (f_1, \dots, f_m)$ y $g = (g_1, \dots, g_p)$ son de clase C^1 , se aplica el resultado anterior a cada componente de $h = g \circ f$ y se concluye que h también es de clase C^1 .

Más generalmente, se dice que una función vectorial $f = (f_1, \dots, f_m)$ es de clase C^p en D si lo es cada una de sus componentes f_1, \dots, f_m (apartado 5.1), o sea, si las funciones f_1, \dots, f_m tienen derivadas parciales continuas en D hasta el orden p . Reiterando el argumento anterior se tiene que si f y g son de clase C^p , $p \geq 2$, la aplicación repetida de la regla de la cadena implica que la función compuesta $h = g \circ f$ también es de clase C^p y permite calcular las derivadas parciales sucesivas de sus componentes hasta el orden p .

Derivada según un vector

Como para las funciones escalares se puede definir la derivada de una función $f : D \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$ según un vector dado $u \in \mathbb{R}^n$. (No imponemos aquí de entrada que u sea un vector unitario, caso para el que reservábamos la denominación de *derivada direccional* —véase la nota 1 del apartado 4.3—, para disponer de un concepto más *analítico* que *geométrico* que no dependa de la norma considerada en \mathbb{R}^n y \mathbb{R}^m .)

Definición 8.4 Sea $f : D \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$, y sean x_0 un punto interior de D y $u \in \mathbb{R}^n$, $u \neq 0$, un vector dado. La *derivada de f según el vector u en x_0* es el límite, si existe

$$\lim_{t \rightarrow 0} \frac{1}{t} [f(x_0 + tu) - f(x_0)] \quad (8.28)$$

y se representa por $D_u f(x_0)$.

Puesto que

$$\lim_{t \rightarrow 0} \frac{1}{t} [f(x_0 + tu) - f(x_0)] = \begin{pmatrix} \lim_{t \rightarrow 0} \frac{1}{t} [f_1(x_0 + tu) - f_1(x_0)] \\ \vdots \\ \lim_{t \rightarrow 0} \frac{1}{t} [f_m(x_0 + tu) - f_m(x_0)] \end{pmatrix}$$

vemos que $D_u f(x_0)$ existe si y sólo si existen las derivadas $D_u f_i(x_0)$, $i = 1, \dots, m$, de las funciones componentes de f y que, si existe,

$$D_u f(x_0) = \begin{pmatrix} D_u f_1(x_0) \\ \vdots \\ D_u f_m(x_0) \end{pmatrix} \quad (8.29)$$

En particular, tomando $u = e_j$, el vector j -ésimo de la base canónica de \mathbb{R}^n , la derivada $D_{e_j}f(x_0)$ es la *derivada parcial*

$$D_j f(x_0) = \begin{pmatrix} \frac{\partial f_1}{\partial x_j}(x_0) \\ \vdots \\ \frac{\partial f_m}{\partial x_j}(x_0) \end{pmatrix}$$

de f respecto a x_j .

Teorema 8.3 Sean $f : D \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$, x_0 un punto interior de D , y f diferenciable en x_0 . Para todo $u \in \mathbb{R}^n$, $u \neq 0$, existe la *derivada de f según el vector u* y viene dada por la fórmula

$$D_u f(x_0) = Df(x_0) \cdot u = \sum_{j=1}^n D_j f(x_0) \cdot u_j \quad (8.30)$$

En efecto, si f es diferenciable en x_0 , también lo son f_1, \dots, f_m y, como vimos en el capítulo 4, existen $D_u f_i(x_0)$, $i = 1, \dots, m$, y se tiene

$$D_u f_i(x_0) = \frac{\partial f_i}{\partial x_1}(x_0) \cdot u_1 + \dots + \frac{\partial f_i}{\partial x_n}(x_0) \cdot u_n$$

Entonces:

$$\begin{aligned} D_u f(x_0) &= \begin{pmatrix} D_u f_1(x_0) \\ \vdots \\ D_u f_m(x_0) \end{pmatrix} = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(x_0) \cdot u_1 + \dots + \frac{\partial f_1}{\partial x_n}(x_0) \cdot u_n \\ \vdots \\ \frac{\partial f_m}{\partial x_1}(x_0) \cdot u_1 + \dots + \frac{\partial f_m}{\partial x_n}(x_0) \cdot u_n \end{pmatrix} = \\ &= \sum_{j=1}^n D_j f(x_0) \cdot u_j = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(x_0) & \dots & \frac{\partial f_1}{\partial x_n}(x_0) \\ \vdots & & \vdots \\ \frac{\partial f_m}{\partial x_1}(x_0) & \dots & \frac{\partial f_m}{\partial x_n}(x_0) \end{pmatrix} \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix} = Df(x_0) \cdot u \end{aligned}$$

El teorema del valor medio

La generalización a funciones vectoriales de los teoremas 4.7 y 4.8 es:

Teorema 8.4 Sean $a, b \in D$, D subconjunto abierto de \mathbb{R}^n , tales que el segmento $[a, b] = \{a + t(b - a) : 0 \leq t \leq 1\}$ está contenido en D . Si

$f : D \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$ es diferenciable en todos los puntos de $[a, b]$, entonces existe $\theta \in (0, 1)$ tal que

$$|f(b) - f(a)| \leq |Df(a + \theta(b-a)) \cdot (b-a)| \quad (8.31)$$

Recuérdese que el teorema 4.8 establece que, bajo esas hipótesis, para cada componente f_i de f existe $\theta_i \in (0, 1)$ tal que

$$f_i(b) - f_i(a) = \nabla f_i(a + \theta_i(b-a)) \cdot (b-a) \quad (8.32)$$

Si el punto intermedio $a + \theta_i(b-a)$ fuese el mismo para todas las componentes, la extensión de la igualdad (8.32) a funciones vectoriales sería inmediata. Pero ello no tiene por qué ser así: sea, por ejemplo, la función $f : \mathbb{R} \rightarrow \mathbb{R}^2$ dada por

$$f(x) = \begin{pmatrix} 1 - \cos x \\ \sin x \end{pmatrix}$$

Se tiene

$$Df(x) \cdot h = h \begin{pmatrix} \sin x \\ \cos x \end{pmatrix} \quad \text{para todo } x \in \mathbb{R}, h \in \mathbb{R}$$

Considerando $a = 0$, $b = 2\pi$ resulta

$$f(2\pi) - f(0) = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \neq 2\pi \begin{pmatrix} \sin 2\pi\theta \\ \cos 2\pi\theta \end{pmatrix} \quad \text{para todo } \theta \in (0, 1)$$

(se tiene $\sin 2\pi\theta = 0$ para $\theta = 1/2$ y $\cos 2\pi\theta = 0$ para $\theta = 1/4$ y $\theta = 3/4$). Así pues, no es posible, en general, obtener para funciones vectoriales la análoga de la igualdad (8.32) y la demostración de (8.31) no es la extensión obvia del resultado (8.32). Veámosla:

Pongamos $z = f(b) - f(a)$. Si $z = 0$, (8.31) es evidente; supongamos entonces $z \neq 0$ y consideremos la función de una variable $\phi : [0, 1] \rightarrow \mathbb{R}$ dada por

$$\phi(t) = \langle z, f(a + t(b-a)) \rangle \quad (8.33)$$

Es una función diferenciable por ser composición de funciones diferenciables (el producto escalar por un elemento fijo de \mathbb{R}^m , $y \mapsto \langle z, y \rangle$, es una aplicación lineal de \mathbb{R}^m en \mathbb{R} y su diferencial es ella misma). Aplicando la regla de la cadena se tiene:

$$\phi'(t) = \langle z, Df(a + t(b-a)) \cdot (b-a) \rangle \quad (8.34)$$

Por el teorema del valor medio para funciones de una variable, existe $\theta \in (0, 1)$ tal que $\phi(1) - \phi(0) = \phi'(\theta) \cdot (1 - 0)$, es decir, utilizando (8.33) y (8.34)

$$\langle z, f(b) - f(a) \rangle = \langle z, Df(a + \theta(b - a)) \cdot (b - a) \rangle \quad (8.35)$$

(obsérvese que (8.35) vale, en realidad, para todo vector $z \in \mathbb{R}^m$). Como $z = f(b) - f(a) \neq 0$,

$$\begin{aligned} |f(b) - f(a)|^2 &= \langle f(b) - f(a), Df(a + \theta(b - a)) \cdot (b - a) \rangle \leq \\ &\leq |f(b) - f(a)| |Df(a + \theta(b - a)) \cdot (b - a)| \end{aligned}$$

de donde resulta (8.31) dividiendo por $|f(b) - f(a)|$.

Corolario 8.1 *Bajo las hipótesis del teorema anterior se tiene*

$$|f(b) - f(a)| \leq \sup_{0 < \theta < 1} \|Df(a + \theta(b - a))\| \cdot |(b - a)| \quad (8.36)$$

$$|f(b) - f(a) - Df(a) \cdot (b - a)| \leq \sup_{0 < \theta < 1} \|Df(a + \theta(b - a)) - Df(a)\| \cdot |(b - a)| \quad (8.37)$$

(8.37) se obtiene de (8.36) aplicado a la función $f - Df(a)$ y teniendo en cuenta que la diferencial de $Df(a)$ es ella misma —en cualquier punto de \mathbb{R}^n — por ser una aplicación lineal. La desigualdad (8.36) se denomina *fórmula de los incrementos finitos* y la (8.37), *fórmula de los incrementos finitos con resto*.

Corolario 8.2 *Si f es diferenciable en un subconjunto convexo C y existe una constante k tal que $\|Df(x)\| \leq k$ para todo $x \in C$, entonces*

$$|f(x) - f(y)| \leq k|x - y| \quad (8.38)$$

para todo par de puntos $x, y \in C$ (o sea, f es lipschitziana en C).

Proposición 8.4 *Sea f diferenciable en un conjunto abierto $D \subset \mathbb{R}^n$ conexo. Es condición necesaria y suficiente para que f sea constante en D el que $Df(x) = 0$ para todo $x \in D$.*

En efecto, basta aplicar a cada componente la proposición 4.2.

Corolario 8.3 *Sean f y g dos funciones diferenciables en un conjunto abierto D conexo tales que $Df(x) = Dg(x)$ para todo $x \in D$. Entonces, existe un vector constante $c = (c_1, \dots, c_m)$ tal que $f(x) = g(x) + c$ para todo $x \in D$, o sea, f y g difieren en una constante.*

Para las funciones de clase C^1 se puede obtener una versión más fuerte de la aproximación (8.4) del incremento de la función mediante su diferencial así como de la proposición 8.3:

Proposición 8.5 Sea D un abierto de \mathbb{R}^n y sea $f: D \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^m$ de clase C^1 en D . Para todo $x_0 \in D$ y todo $\varepsilon > 0$ existe $\delta > 0$ tal que $B_\delta(x_0) \subseteq D$,

$$|f(x) - f(y) - Df(x_0) \cdot (x - y)| \leq \varepsilon |x - y| \quad \text{para todo } x, y \in B_\delta(x_0) \quad (8.39)$$

y, en consecuencia,

$$|f(x) - f(y)| \leq (\|Df(x_0)\| + \varepsilon) |x - y| \quad \text{para todo } x, y \in B_\delta(x_0) \quad (8.40)$$

Demostración: Consideremos, como antes, $g = f - Df(x_0)$. Se tiene $Dg(x_0) = Df(x_0) - Df(x_0) = 0$ (la aplicación —matriz— cero). Puesto que

$$Dg: D \rightarrow \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$$

es continua y $\|Dg(x_0)\| = 0$, existe $B_\delta(x_0) \subseteq D$ tal que, si $x \in B_\delta(x_0)$, $\|Dg(x)\| < \varepsilon$. Por el corolario 8.2, se tendrá $|g(x) - g(y)| \leq \varepsilon |x - y|$, es decir, $|f(x) - f(y) - Df(x_0) \cdot (x - y)| \leq \varepsilon |x - y|$ para todo $x, y \in B_\delta(x_0)$. (8.40) resulta de la desigualdad triangular aplicada a

$$|f(x) - f(y)| = |[f(x) - Df(x_0) \cdot x] - [f(y) - Df(x_0) \cdot y] + Df(x_0) \cdot (y - x)|$$

8.3 Curvas en \mathbb{R}^n

Una curva γ en \mathbb{R}^n es la imagen de un intervalo $I \subseteq \mathbb{R}$ por una función continua $\phi: I \rightarrow \mathbb{R}^n$: $\gamma = \{x \in \mathbb{R}^n; x = \phi(t) \text{ para algún } t \in I\}$ (en el apartado 7.7 utilizamos el término *arco* para el caso en que I es compacto). Con cierto abuso de lenguaje llamaremos en ocasiones curva a la propia función ϕ de la cual es imagen.

Ejemplos

1.- El ejemplo más simple de curva es (valga el juego de palabras) una recta:

$$x(t) = x_0 + tu, \quad t \in \mathbb{R} \quad (8.41)$$

donde $x_0, u \in \mathbb{R}^n$, $u \neq 0$, se suponen dados (es la recta que pasa por x_0 y tiene a u por vector dirección). Desplegando (8.41) en coordenadas se tienen las expresiones de las funciones componentes:

$$\begin{cases} x_1(t) = x_1^0 + tu_1 \\ \vdots \\ x_n(t) = x_n^0 + tu_n \end{cases} \quad (8.42)$$

Si $t \in [0, 1]$, se tiene el segmento $[x_0, y_0]$, con $y_0 = x_0 + u$.

2.- En \mathbb{R}^2 , la curva definida por

$$\begin{cases} x(t) = a \cos t \\ y(t) = b \sin t \end{cases}, \quad 0 \leq t \leq 2\pi \quad (8.43)$$

es una *elipse* de semiejes a y b . Si $a = b$, se trata de la *circunferencia* de centro el origen y radio a .

3.- La curva en \mathbb{R}^2 definida por

$$\begin{cases} x(t) = a(t - \sin t) \\ y(t) = a(1 - \cos t) \end{cases}, \quad t \in \mathbb{R} \quad (8.44)$$

se llama *cicloide* (figura 8.1).

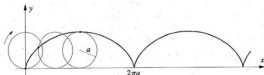


Figura 8.1

4.- En \mathbb{R}^3 , la curva definida por

$$\begin{cases} x(t) = a \cos t \\ y(t) = a \sin t \\ z(t) = ht \end{cases}, \quad t \in \mathbb{R} \quad (8.45)$$

es una *hélice*. Si $h > 0$, se dice que la hélice es *dextrógira* y si $h < 0$, *levógira*. El número $2\pi|h|$ es el *paso* de la hélice (figura 8.2).

Insistamos en que llamamos aquí *curva* a la imagen de $\phi : I \rightarrow \mathbb{R}^n$, no a su gráfica, que es el conjunto $\{(t, x_1, \dots, x_n); x_1 = \phi_1(t), \dots, x_n = \phi_n(t)\}$ contenido en \mathbb{R}^{n+1} . Por ejemplo, la gráfica de la función (8.43) es precisamente una hélice en \mathbb{R}^3 como la del ejemplo 4.

El lector reconocerá sin duda, en los anteriores, ejemplos de curvas dadas por sus *ecuaciones paramétricas* (el *parámetro* es la variable t) y, como hemos recordado en el primer capítulo, conoce otras formas de representar curvas: la gráfica en el plano de una función $y = f(x)$ es, en general, una curva; cualquier

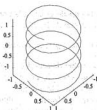


Figura 8.2

curva dada en forma explícita $y = f(x)$ se puede describir paramétricamente introduciendo, simplemente, la variable x como parámetro y poniendo

$$\begin{cases} x = t \\ y = f(t) \end{cases} \quad (8.46)$$

Recíprocamente, dada una curva $x_i(t) = \phi_i(t)$, $i = 1, \dots, n$, si una de las funciones ϕ_i es invertible, por ejemplo ϕ_k , y su inversa $t = \psi(x_k)$ está definida y es continua en un cierto intervalo J del eje x_k , entonces las ecuaciones

$$x_i = \phi_i(\psi(x_k)) = \phi(x_k), \quad x_k \in J, \quad i = 1, \dots, n, \quad i \neq k \quad (8.47)$$

constituyen una representación explícita, no paramétrica de dicha curva (cuando es posible ese paso se dice que se ha *eliminado el parámetro*). No siempre se puede conseguir una representación explícita global de una curva; así ocurre, por ejemplo, con la circunferencia unidad

$$x(t) = \cos t, \quad y(t) = \sin t, \quad 0 \leq t \leq 2\pi \quad (8.48)$$

si bien la semicircunferencia superior ($0 \leq t \leq \pi$) y la inferior ($\pi \leq t \leq 2\pi$) tienen, respectivamente, las representaciones explícitas

$$y = \sqrt{1 - x^2}, \quad y = -\sqrt{1 - x^2}$$

Otra forma de representación no paramétrica de curvas es la implícita: en el plano, una curva es, bajo ciertas hipótesis, la gráfica de una ecuación en dos variables

$$F(x, y) = 0 \quad (8.49)$$

Así, la circunferencia (8.48) es la gráfica de la ecuación $x^2 + y^2 = 1$. En el caso de curvas en \mathbb{R}^n , la representación implícita es un sistema de $n-1$ ecuaciones en las variables x_1, \dots, x_n .

Por otro lado, una curva puede venir dada por diferentes ecuaciones paramétricas; por ejemplo, la imagen de (8.48) es la misma que la de la función

$$x(t) = \cos 2\pi t, y(t) = \sin 2\pi t, 0 \leq t \leq 1 \quad (8.50)$$

a saber, la circunferencia unidad de ecuación implícita $x^2 + y^2 = 1$.

La representación paramétrica, o, lo que es lo mismo, la utilización de funciones vectoriales de una variable, permite en muchas ocasiones describir fácilmente curvas para las que podría ser muy difícil o imposible encontrar representaciones explícitas o implícitas no paramétricas (en el caso de la cicloide, por ejemplo, la eliminación del parámetro llevaría a una relación complicada entre x e y). Pero, sobre todo, su interés radica en que una curva aparece no solamente como el conjunto de valores de una función, no únicamente como un objeto geométrico en el espacio de las variables x_1, \dots, x_n , sino como una trayectoria de puntos, los valores $x = \phi(t)$, que se van recorriendo según el orden correspondiente a valores crecientes del parámetro t ; queda así determinada una *orientación* en la curva. Por ejemplo,

$$x(t) = \cos t, y(t) = \sin t, 0 \leq t \leq 2\pi$$

es la circunferencia $x^2 + y^2 = 1$ recorrida en sentido positivo, o sea, el contrario a las agujas del reloj, mientras que

$$x(t) = \cos t, y(t) = -\sin t, 0 \leq t \leq 2\pi \quad (8.51)$$

tiene también por imagen los puntos de \mathbb{R}^2 que satisfacen $x^2 + y^2 = 1$ pero ahora la circunferencia unidad se recorre en sentido negativo (se puede dar, aunque no entraremos aquí en ello, una definición rigurosa de funciones $\phi: I \rightarrow \mathbb{R}^n$, $\phi^*: I^* \rightarrow \mathbb{R}^n$ *equivalentes*, que son aquellas que tienen la misma imagen recorrida en el mismo sentido, o sea, que definen la misma curva orientada; (8.48) y (8.51) no son equivalentes aunque tengan la misma imagen mientras que (8.48) y (8.50) sí lo son).

El parámetro t representa, en muchas aplicaciones, el tiempo, y, así, la idea intuitiva de curva en el sentido que aquí exponemos es la de trayectoria de un punto que se mueve durante un intervalo de tiempo $t \in I$. Este punto de vista es muy útil en el estudio del movimiento y en otras aplicaciones no directamente físicas pero que, de algún modo, se pueden asimilar por analogía con algún fenómeno físico o geométrico previamente estudiado.

Una curva $x = \phi(t)$, $t \in [a, b]$ se dice que es *cerrada* si $\phi(a) = \phi(b)$, o sea, si coinciden los extremos inicial y final. Una curva $x = \phi(t)$, $t \in I$ se dice que es *simple* (o que es una curva de Jordan) si $\phi(t_1) \neq \phi(t_2)$ siempre que $t_1, t_2 \in I$ verifiquen $t_1 \neq t_2$, o sea, si sólo se pasa una vez por cada uno de los puntos de la imagen γ de ϕ (para una curva cerrada se entiende que esa

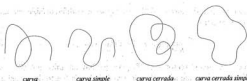


Figura 8.3

condición es para t_1, t_2 tales que $|t_1 - t_2| < b - a$; una curva simple no se corta a sí misma; nótese que toda curva que admite una representación explícita no paramétrica es simple, pero el recíproco no es cierto como muestra el caso de una circunferencia; figura 8.3).

En realidad, la idea intuitiva de curva se corresponde más bien con funciones derivables que permitan dar sentido a la tangente en cada punto (salvo, acaso, en algunos puntos excepcionales). Se pueden, en efecto, contruir "curvas" en el sentido de la definición original, o sea, imágenes de funciones continuas, que llenan, por ejemplo, todo un cuadrado del plano.

Curvas diferenciables

Por curva diferenciable entendemos, naturalmente, una función $\phi : I \rightarrow \mathbb{R}^n$ diferenciable, lo que, por tratarse de funciones de una variable, significa que sus componentes son funciones derivables en todo $t \in I$. Una curva de clase C^p es una función $\phi : I \rightarrow \mathbb{R}^n$ cuyas componentes $\phi_i : I \rightarrow \mathbb{R}$ tienen derivadas continuas en todo $t \in I$ hasta el orden p .

La diferenciable en $t_0 \in I$ significa que hay un vector que denotamos por $\phi'(t_0)$ tal que

$$\lim_{h \rightarrow 0} \frac{1}{|h|} [\phi(t_0 + h) - \phi(t_0) - h\phi'(t_0)] = 0 \quad (8.52)$$

o, equivalentemente,

$$\lim_{t \rightarrow t_0} \frac{1}{|t - t_0|} [\phi(t) - \phi(t_0) - (t - t_0)\phi'(t_0)] = 0 \quad (8.53)$$

o también

$$\lim_{t \rightarrow t_0} \left| \frac{\phi(t) - \phi(t_0)}{t - t_0} - \phi'(t_0) \right| = 0 \quad (8.54)$$

Obsérvese (cf. el apartado 3.1) que, supuesta ϕ diferenciable, entonces es continua y se tiene

$$\lim_{h \rightarrow 0} [\phi(t_0 + h) - \phi(t_0) - ha] = 0 \quad (8.55)$$

cualquiera que sea el vector $a \in \mathbb{R}^n$, pero que hay un único $a \in \mathbb{R}^n$, a saber, $a = \phi'(t_0)$, para el que se verifica la propiedad más fuerte de aproximación (8.52). Esto quiere decir que, supuesto que $\phi'(t_0) \neq 0$, la recta

$$X(t) = \phi(t_0) + (t - t_0)\phi'(t_0) \quad (8.56)$$

es, entre las rectas $\phi(t_0) + (t - t_0)a$ que pasan por $\phi(t_0)$, la que mejor aproxima los valores de la función $\phi(t)$ en las proximidades de $\phi(t_0)$. La llamamos, por ello, **recta tangente** a la curva $\phi(t)$ en el punto $\phi(t_0)$ y al vector $\phi'(t_0)$, **vector tangente** a la curva en $\phi(t_0)$; lo visualizamos trasladándolo desde el origen a $\phi(t_0)$, o sea, con base en $\phi(t_0)$ y extremo en $\phi(t_0) + \phi'(t_0)$ (figura 8.4).

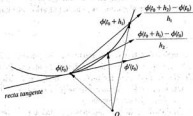


Figura 8.4

Nótese que la recta tangente y el vector tangente sólo están definidos si $\phi'(t_0) \neq 0$ (en la cicloide, $\phi'(2k\pi) = (0, 0)$ y se ve en la figura 8.1 que no está definida la tangente en los puntos $\phi(2k\pi) = (2k\pi a, 0)$). Por otro lado, si una curva no es simple puede tener distintas tangentes en alguno de los puntos por los que se pasa más de una vez, como se aprecia en la figura 8.5 para la *cicloide prolata* definida por las ecuaciones

$$x(t) = 2t - \pi \sin t, \quad y(t) = 2 - \pi \cos t$$

en el punto $(x(-\pi/2), y(-\pi/2)) = (x(\pi/2), y(\pi/2)) = (0, 2)$ en que se corta a sí misma.

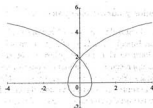


Figura 8.5

Aplicuese lo anterior como ejercicio a la gráfica de una función $y = f(x)$ de una variable, o sea a la curva $x = t$, $y = f(t)$, comparando con las observaciones hechas en el apartado 3.1.

Si $\phi(t)$ describe el movimiento de una partícula en el espacio \mathbb{R}^n , el cociente

$$\frac{\phi_i(t) - \phi_i(t_0)}{t - t_0}, \quad i = 1, \dots, n$$

representa la velocidad media a la que la proyección de la partícula sobre el eje x_i ha recorrido el espacio $\phi_i(t) - \phi_i(t_0)$. Parece, pues, natural llamar **vector velocidad (instantánea)** al vector $\phi'(t_0)$. El módulo o norma de este vector es la **velocidad (instantánea)** de la partícula en el instante t_0 . Esta interpretación *dinámica* de una curva y de su vector tangente (nótese que éste señala el sentido de avance sobre $\phi(t)$ al crecer t) es útil en otros campos diferentes a la física. Por ejemplo, para la curva

$$x(t) = \cos t, \quad y(t) = \sin t, \quad 0 \leq t \leq 2\pi$$

el vector velocidad es $v(t) = (x'(t), y'(t)) = (-\sin t, \cos t)$ y la velocidad (escalar) $|v(t)| = \sqrt{(-\sin t)^2 + \cos^2 t} = 1$.

Se define como **vector aceleración** en $t = t_0$ del movimiento definido por $\phi(t)$ a la derivada en t_0 , si existe, de la función $t \mapsto v(t) = \phi'(t)$, o sea, a $\phi''(t_0)$. En el ejemplo anterior, $\phi''(t) = (-\cos t, -\sin t)$. Obsérvese que la aceleración no es nula porque el vector velocidad, aun siendo de módulo constante, varía en dirección con el tiempo. Hay dos posibles definiciones de **aceleración escalar**: bien como

$$\frac{d}{dt}|v(t)|$$

bien como $|\phi''(t)|$, que, como regla general, no coinciden: en este ejemplo,

$$\frac{d}{dt}|v(t)| = 0 \quad \text{y} \quad |\phi''(t)| = 1$$

Campos vectoriales y curvas integrales

Hemos definido en el capítulo 5 un *campo vectorial* sobre un conjunto abierto $\Omega \subseteq \mathbb{R}^n$ como una función $f: \Omega \rightarrow \mathbb{R}^n$ y lo visualizábamos dibujando el vector $f(x)$ con base en el punto x (véase la figura 5.1). En Física, los campos vectoriales aparecen principalmente como *campos de fuerzas* o como *campos de velocidades*. En el primer caso, el vector $f(x)$ se interpreta como una fuerza que actúa sobre una partícula colocada en x . Por ejemplo, la ley de gravitación de Newton establece que un cuerpo de masa M ejerce una fuerza sobre un cuerpo de masa m que, en magnitud, vale gMm/r^2 , donde g es una constante (la constante de gravitación) y r la distancia entre los centros de gravedad, y está dirigida de m a M . Si M es mucho mayor que m (caso del sol respecto a los planetas) se puede suponer que M está situada en un origen fijo de coordenadas de \mathbb{R}^3 y entonces la fuerza de atracción que ejerce sobre una masa m dada situada en $x = (x_1, x_2, x_3)$ es

$$f(x) = -K \frac{x}{|x|^3} \quad (8.57)$$

o sea

$$f(x) = (f_1(x), f_2(x), f_3(x)) = \left(-K \frac{x_1}{|x|^3}, -K \frac{x_2}{|x|^3}, -K \frac{x_3}{|x|^3} \right) \quad (8.58)$$

siendo K una constante (obsérvese que el vector $f(x)$ está dirigido hacia el origen; esbócese una figura en el plano).

En el caso de un campo de velocidades, $f(x)$ se interpreta como el vector velocidad en el punto x de una partícula que se mueve de acuerdo a la ley física relevante en el fenómeno bajo estudio. Por ejemplo, supongamos en el plano una placa circular (un antiguo disco de vinilo, por ejemplo) que gira en sentido positivo alrededor del origen con una frecuencia de ω vueltas por minuto. Entonces, el punto de la placa que está en un instante dado en (x, y) se mueve con un vector velocidad $f(x, y) = (-2\pi\omega y, 2\pi\omega x)$. En efecto, $\phi(t) = (r \cos 2\pi\omega t, r \sin 2\pi\omega t)$ describe el movimiento de giro, en sentido positivo con frecuencia ω , de un punto que está a una distancia r del origen. Entonces, $\phi'(t) = (-2\pi\omega r \sin 2\pi\omega t, 2\pi\omega r \cos 2\pi\omega t) = (-2\pi\omega y, 2\pi\omega x)$ (dibújese también aquí un esbozo de este campo de velocidades).

Muchos problemas dinámicos, o sea, problemas que surgen en el estudio de fenómenos que evolucionan en el tiempo, se ajustan al siguiente modelo: La ley física (o biológica, económica, etcétera) que rige el fenómeno bajo estudio define un campo vectorial $f(x)$ y se plantea la cuestión de determinar las curvas $\phi(t)$ (llamadas *curvas integrales* del campo) que describen la evolución de dicho fenómeno con la condición de que, en cada uno de sus puntos, el vector

tangente ha de ser precisamente el vector que asigna el campo, o sea, tales que

$$\phi'(t) = f(\phi(t)) \quad (8.59)$$

para todo t del intervalo de definición de $\phi(t)$ (el campo $f(x)$ es, pues, un campo de velocidades y sus curvas integrales son las curvas que, partiendo de un estado inicial $x_0 = \phi(0)$, van "buscando" ser en cada punto tangentes al campo; figura 8.6).

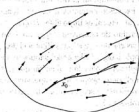


Figura 8.6

Analíticamente, el problema de la determinación de las curvas integrales de un campo vectorial $f(x)$ es de la resolución del sistema de ecuaciones diferenciales:

$$\begin{cases} x_1' = f_1(x_1, \dots, x_n) \\ \vdots \\ x_n' = f_n(x_1, \dots, x_n) \end{cases} \quad (8.60)$$

o, en forma de ecuación diferencial vectorial

$$x' = f(x) \quad (8.61)$$

El resultado fundamental de la teoría de ecuaciones diferenciales establece que si f es de clase C^1 en un abierto $\Omega \subseteq \mathbb{R}^n$, entonces, para cada $x_0 \in \Omega$, existe una única solución $\phi(t)$ de $x' = f(x)$, o sea, tal que $\phi'(t) = f(\phi(t))$ para todo t de un cierto intervalo (α, ω) , que satisface además la condición inicial $\phi(0) = x_0$. Esto se corresponde con lo que cabía prever desde un punto de vista determinista, a saber, que si se conoce el estado inicial del sistema físico que se estudia, expresado por el vector x_0 , y la ley que rige su evolución, expresada en forma de ecuación diferencial, entonces queda unívocamente determinada la evolución subsiguiente de dicho sistema. Las curvas que definen las soluciones $\phi(t)$ de (8.60) se llaman trayectorias del sistema. Unas soluciones muy especiales que puede tener un sistema de ecuaciones diferenciales

son las soluciones constantes o de equilibrio; compruébese que $\phi(t) = \bar{x}$ para todo t si y sólo si $\bar{x} \in \Omega$ es tal que $f(\bar{x}) = 0$. Las curvas que definen son curvas constantes, es decir, puntos de Ω que reciben el nombre de puntos de equilibrio del sistema por el significado dinámico que tienen: si el sistema está en \bar{x} en el instante inicial $t = 0$, permanecerá para siempre (y siempre estuvo) en \bar{x} (obsérvese que en dichos puntos el vector del campo es nulo y no está definido vector tangente). Los puntos de equilibrio, si existen, suelen desempeñar un papel relevante en la dinámica de conjunto del sistema.

A diferencia de los casos escalares a los que hemos hecho referencia en capítulos anteriores, en general es imposible resolver un sistema de ecuaciones diferenciales en el sentido de encontrar explícitamente fórmulas que nos den sus soluciones. La teoría de ecuaciones diferenciales permite, sin embargo, obtener sobre las propiedades de dichas soluciones una amplia información que en muchos casos es suficiente para la aplicación que nos interese, en particular cuando el campo $f(x)$ tiene propiedades especiales, como, por ejemplo, si se trata de un campo gradiente, o sea, si es tal que existe una función $F: \Omega \rightarrow \mathbb{R}$ tal que $\nabla F = f$ (problema 7).

Consideremos ahora el caso de una partícula que se mueve por la acción de un campo de fuerzas $f(x)$, como, por ejemplo, un planeta en el campo gravitatorio creado por el sol (o una partícula cargada en un campo eléctrico, o magnético, etcétera). La ley que rige los posibles movimientos de la partícula es la segunda ley de Newton, de acuerdo con la cual, si $\phi(t)$ denota el vector posición de aquella y la función $\phi(t)$ es al menos de clase C^2 , se verifica en todo instante t la relación

$$f(\phi(t)) = m\phi''(t) \quad (8.62)$$

donde m es la masa de la partícula. La búsqueda de los movimientos $\phi(t)$ que responden a esta ley conduce a la resolución de un sistema de n ($n = 1, 2$ ó 3 en las aplicaciones usuales) ecuaciones diferenciales de segundo orden (por intervenir la derivadas segundas de las funciones incógnitas):

$$\begin{cases} x_1'' = \frac{1}{m} f_1(x_1, \dots, x_n) \\ \vdots \\ x_n'' = \frac{1}{m} f_n(x_1, \dots, x_n) \end{cases} \quad (8.63)$$

En cada situación concreta, f tiene una forma determinada; en el caso del campo gravitatorio, está dada por (8.57). La manera de tratar el sistema (8.63) consiste en reducirlo a uno del tipo (8.60) introduciendo como nueva variable suplementaria el vector velocidad

$$v(t) = (\phi_1'(t), \dots, \phi_n'(t))$$

de tal modo que el sistema (8.63) resulta equivalente (de las soluciones de uno se deducen las del otro) al sistema de $2n$ ecuaciones de primer orden

$$\begin{cases} x'_1 = v_1 \\ \vdots \\ x'_n = v_n \\ v'_1 = \frac{1}{m} f_1(x_1, \dots, x_n) \\ \vdots \\ v'_n = \frac{1}{m} f_n(x_1, \dots, x_n) \end{cases} \quad (8.64)$$

al cual se aplica la teoría general elaborada para los sistemas en la forma canónica (8.60). En las trayectorias de este sistema, que son curvas en \mathbb{R}^{2n} , queda reflejada la evolución conjunta de los vectores posición y velocidad de los movimientos regidos por la ley (8.62) (problema 8).

8.4 Notas y complementos

1. Diferenciales de orden superior

Sea $f: D \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^m$ y supongamos que existe $Df(x)$ para todo $x \in D$. Se tiene entonces definida una aplicación Df :

$$D \ni x \xrightarrow{Df} Df(x) \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m) \approx \mathbb{R}^{m \times n} \quad (8.65)$$

Si a su vez esta aplicación es diferenciable en un punto $x_0 \in D$ (respectivamente en todo D), se dice que f es *dos veces diferenciable* en x_0 (resp. en D) y la diferencial de Df en x_0 se llama *diferencial segunda* de f en x_0 y se representa por $D^2f(x_0)$ o $f''(x_0)$. Es un elemento del espacio $\mathcal{L}(\mathbb{R}^n, \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m))$. (Más generalmente, sin exigir que f sea diferenciable en todo D , se dice que f es dos veces diferenciable en $x_0 \in D$ si f es diferenciable en un entorno U de x_0 y la aplicación $Df: U \rightarrow \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ es diferenciable en x_0 .)

No es fácil la interpretación directa de los elementos del espacio $\mathcal{L}(\mathbb{R}^n, \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m))$, pero en álgebra lineal se ve que se pueden identificar con las aplicaciones bilineales de \mathbb{R}^n en \mathbb{R}^m . Recordemos que una aplicación

$$(x, y) \ni \mathbb{R}^n \times \mathbb{R}^n \xrightarrow{B} B(x, y) \in \mathbb{R}^m$$

es bilineal si es lineal en x para cada y fijo y lineal en y para cada x fijo, o sea, si cualesquiera que sean $x, y, z \in \mathbb{R}^n$ y $\alpha, \beta \in \mathbb{R}$ se tiene:

$$\begin{aligned} B(\alpha x + \beta z, y) &= \alpha B(x, y) + \beta B(z, y) \\ B(x, \alpha y + \beta z) &= \alpha B(x, y) + \beta B(x, z) \end{aligned}$$

Una aplicación bilineal se dice que es *simétrica* si $B(x, y) = B(y, x)$ para todo par de puntos $x, y \in \mathbb{R}^n$. El conjunto de aplicaciones bilineales de \mathbb{R}^n en \mathbb{R}^m es un espacio vectorial que se denota por $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^n; \mathbb{R}^m)$ o $\mathcal{L}_2(\mathbb{R}^n, \mathbb{R}^m)$. Recordemos también cómo se logra la identificación mencionada. Sea $H \in \mathcal{L}(\mathbb{R}^n, \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m))$; tomemos un $x \in \mathbb{R}^n$ cualquiera y pongamos $L_x = H(x)$. Entonces, $L_x \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ y $B(x, y) \stackrel{\text{def}}{=} L_x(y) \in \mathbb{R}^m$ para todo $y \in \mathbb{R}^n$. La aplicación $(x, y) \mapsto B(x, y)$ así definida se comprueba fácilmente que es bilineal y, por tanto, a todo elemento de $\mathcal{L}(\mathbb{R}^n, \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m))$ queda asociado una aplicación bilineal $B \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^n; \mathbb{R}^m)$. Recíprocamente, toda aplicación bilineal $B \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^n; \mathbb{R}^m)$ es la imagen por esta correspondencia de un elemento de $\mathcal{L}(\mathbb{R}^n, \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m))$; en efecto, dada $B(x, y)$, entonces, fijado $x \in \mathbb{R}^n$, la aplicación $H(x) \stackrel{\text{def}}{=} B(x, \cdot)$ pertenece a $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$, y, por tanto, $H: x \mapsto B(x, \cdot)$ pertenece a $\mathcal{L}(\mathbb{R}^n, \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m))$. Se comprueba que esta correspondencia es lineal y biunívoca y, en consecuencia, se pueden identificar ambos espacios (de hecho, no sólo como espacios vectoriales sino también en tanto que espacios normados: lo mismo que para las aplicaciones lineales, se puede definir una norma en el espacio $\mathcal{L}_2(\mathbb{R}^n, \mathbb{R}^m)$ de aplicaciones bilineales por

$$\|B\| = \max \{|B(x, y)|; |x|, |y| \leq 1\}$$

y se comprueba que, en la correspondencia anterior, $\|B\| = \|H\|$, o sea, que se trata de una *isometría* entre $\mathcal{L}_2(\mathbb{R}^n, \mathbb{R}^m)$ y $\mathcal{L}(\mathbb{R}^n, \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m))$).

Concretando esto a lo que aquí nos interesa, si $D^2f(x_0) \in \mathcal{L}(\mathbb{R}^n, \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m))$ existe, se identifica con la aplicación bilineal

$$(h, k) \mapsto (D^2f(x_0) \cdot h) \cdot k \stackrel{\text{def}}{=} D^2f(x_0) \cdot (h, k) \quad (8.66)$$

(utilizamos, como en el apartado 8.1, la notación $L \cdot h$ para indicar la actuación de una aplicación lineal sobre un vector para no sobrecargar la escritura de paréntesis y, así, la interpretación de (8.66) es la siguiente: h y k son vectores de \mathbb{R}^n ; como $D^2f(x_0)$ es una aplicación lineal $\mathbb{R}^n \rightarrow \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$, su valor en $h \in \mathbb{R}^n$ es un elemento $D^2f(x_0) \cdot h \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$, es decir, $D^2f(x_0) \cdot h$ es una aplicación lineal $\mathbb{R}^n \rightarrow \mathbb{R}^m$; y su valor en $k \in \mathbb{R}^n$ se denota por $(D^2f(x_0) \cdot h) \cdot k$, que es un elemento de \mathbb{R}^m). Se puede demostrar que si f es dos veces diferenciable en x_0 , $D^2f(x_0)$ es una aplicación bilineal simétrica, es decir

$$D^2f(x_0) \cdot (h, k) = D^2f(x_0) \cdot (k, h) \text{ para todo par } h, k \in \mathbb{R}^n \quad (8.67)$$

(véase [3] o [5], obras a las se remite al lector interesado en profundizar en las cuestiones esbozadas en esta nota).

Las aplicaciones bilineales de \mathbb{R}^n en \mathbb{R} , llamadas *formas bilineales en \mathbb{R}^n* , están representadas por matrices $n \times n$: Dados $x = (x_1, \dots, x_n)$, $y = (y_1, \dots, y_n)$ se tiene:

$$\begin{aligned} B(x, y) &= B\left(\sum_{i=1}^n x_i e_i, \sum_{j=1}^n y_j e_j\right) = \sum_{i=1}^n \sum_{j=1}^n x_i y_j B(e_i, e_j) = \\ &= \sum_{i=1}^n \sum_{j=1}^n b_{ij} x_i y_j = x^T B y \end{aligned}$$

donde llamamos también B a la matriz de elementos $b_{ij} = B(e_i, e_j)$. Recíprocamente, dada una matriz B $n \times n$, es inmediato comprobar que $(x, y) \mapsto x^T B y$ es una forma bilineal en \mathbb{R}^n (Cuando en una forma bilineal hacemos $y = x$ se obtiene una *forma cuadrática* $B(x, x)$ en \mathbb{R}^n .)

Si $f : D \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ es dos veces diferenciable en $x_0 \in D$ ¿cómo será $D^2 f(x_0)$ en tanto que forma bilineal en \mathbb{R}^n ? Sea $A = (a_{ij})$ la matriz que representa a $D^2 f(x_0)$; entonces, $a_{ij} = D^2 f(x_0) \cdot (e_i, e_j)$ y

$$D^2 f(x_0) \cdot (h, k) = \sum_{i=1}^n \sum_{j=1}^n a_{ij} h_i k_j$$

Además, A es simétrica por (8.67). Por hipótesis, f es diferenciable en todo punto x de un cierto entorno U de x_0 y, como sabemos (apartado 3.3), $Df(x) : \mathbb{R}^n \rightarrow \mathbb{R}$ está dada por

$$Df(x) \cdot (k_1, \dots, k_n) = \frac{\partial f}{\partial x_1}(x) k_1 + \dots + \frac{\partial f}{\partial x_n}(x) k_n \quad (8.68)$$

(en el apartado 3.3 representábamos este valor por $df(x; h)$) y, a su vez, $Df : U \subseteq \mathbb{R}^n \rightarrow \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ es diferenciable en x_0 . Entonces

$$\begin{aligned} (D^2 f(x_0) \cdot h) \cdot k &= (\text{por (8.30)}) = \left(\lim_{t \rightarrow 0} \frac{1}{t} [Df(x_0 + th) - Df(x_0)] \right) \cdot k = \\ &= \lim_{t \rightarrow 0} \frac{1}{t} [Df(x_0 + th) \cdot k - Df(x_0) \cdot k] = \\ &= \lim_{t \rightarrow 0} \frac{1}{t} \left[\sum_{i=1}^n \frac{\partial f}{\partial x_i}(x_0 + th) \cdot k_i - \sum_{i=1}^n \frac{\partial f}{\partial x_i}(x_0) \cdot k_i \right] = \\ &= \sum_{i=1}^n \lim_{t \rightarrow 0} \frac{1}{t} \left[\frac{\partial f}{\partial x_i}(x_0 + th) - \frac{\partial f}{\partial x_i}(x_0) \right] \cdot k_i \end{aligned}$$

Tomando los vectores de la base canónica $h = e_j$, $k = e_i$, resulta

$$a_{ji} = \frac{\partial^2 f}{\partial x_i \partial x_j}(x_0) = (\text{por la simetría de } A) = a_{ij}$$

de modo que, como cabía esperar por los comentarios que hicimos en el capítulo 5, la matriz que da la diferencial segunda de f no es otra que su *matriz hessiana*. (De la demostración anterior vemos que si f es dos veces diferenciable, entonces f posee derivadas parciales segundas; se verifica además que

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(x_0) = \frac{\partial^2 f}{\partial x_j \partial x_i}(x_0) \quad (8.69)$$

por el resultado (8.67), es decir la igualdad de las derivadas segundas cruzadas. Las hipótesis del teorema de Schwarz y sus variantes que aseguran la existencia de las derivadas segundas de f y la igualdad (8.69) son condiciones suficientes para que f sea dos veces diferenciable.)

Si la diferencial segunda existe en todo punto de D , define una aplicación

$$D \rightarrow \mathcal{L}(\mathbb{R}^n, \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)) \approx \mathcal{L}_2(\mathbb{R}^n, \mathbb{R}^m)$$

cuya diferencial en x_0 , si existe, es la *diferencial tercera* de f en x_0 . Y, en general, se define por recurrencia la *diferencial p -ésima*, $D^p f(x_0)$, como la diferencial de la diferencial de orden $p-1$. $D^p f(x_0)$ se identifica con un elemento del espacio $\mathcal{L}_p(\mathbb{R}^n, \mathbb{R}^m)$ de aplicaciones p -lineales de \mathbb{R}^n en \mathbb{R}^m :

$$\mathbb{R}^n \times \underbrace{\cdots}_p \times \mathbb{R}^n \ni (h^1, \dots, h^p) \xrightarrow{B} D^p f(x_0) \cdot (h^1, \dots, h^p) \in \mathbb{R}^m$$

que se definen de manera análoga a las aplicaciones bilineales.

2. Generalización a espacios normados

Las nociones del cálculo diferencial en los espacios \mathbb{R}^n se extienden casi literalmente, con algunas modificaciones, al caso de espacios normados:

Sean E y F dos espacios normados y sea D un subconjunto abierto de E . Se dice que una función $f: D \rightarrow F$ es **diferenciable Fréchet** en el punto $x_0 \in D$ si existe una aplicación lineal y continua $L: E \rightarrow F$ tal que

$$\lim_{h \rightarrow 0} \frac{1}{\|h\|} [f(x_0 + h) - f(x_0) - L(h)] = 0 \quad (8.70)$$

Hay en esta definición una diferencia respecto a la definición 8.1: en dimensión finita, toda aplicación lineal $L: \mathbb{R}^n \rightarrow \mathbb{R}^m$ es continua, como vimos en el apartado 7.6, pero esto no tiene por que ser así en espacios normados de dimensión infinita. Para mantener el objetivo esencial del cálculo diferencial, a saber, la *aproximación local* de una aplicación arbitraria por una aplicación

lineal, ha de imponerse de entrada que ésta sea continua. El conjunto de aplicaciones lineales continuas $L: E \rightarrow F$ se representa por $\mathcal{L}(E, F)$. Se puede demostrar la siguiente caracterización:

Dada una aplicación lineal $L: E \rightarrow F$, las tres condiciones siguientes son equivalentes:

- (i) L es continua en todo punto de E .
- (ii) L es continua en el origen $O \in E$.
- (iii) $|L(x)|$ está acotada en la bola unidad $|x| \leq 1$.

De (iii) se deduce que $|L(x)| \leq k|x|$ para un cierto $k \in \mathbb{R}$; la menor de esas cotas k es $\|L\| = \max\{|L(x)|; |x| \leq 1\}$, se llama *norma* de L y, efectivamente, $L \mapsto \|L\|$ es una norma en $\mathcal{L}(E, F)$ (es un espacio de Banach si F lo es).

Como en el caso de los espacios \mathbb{R}^n , se demuestra que hay a lo más una aplicación $L \in \mathcal{L}(E, F)$ que satisface (8.70) y también, prácticamente con las mismas demostraciones, se demuestran todos los resultados expuestos en los apartados anteriores y en la nota precedente.

También se define de la misma manera la derivada según un vector: Sea $f: D \rightarrow F$, y sean $x_0 \in D$ y $u \in E$, $u \neq 0$, un vector dado. La *derivada de f según el vector u en x_0* es el límite, si existe

$$\lim_{t \rightarrow 0} \frac{1}{t} [f(x_0 + tu) - f(x_0)]$$

y se representa por $d_u f(x_0)$ o $df(x_0, u)$. Si $df(x_0, u)$ existe para todo $u \in E$ y la aplicación de E en F dada por $u \mapsto df(x_0, u)$ es lineal y continua, se denomina **diferencial de Gateaux** de f en x_0 y se dice que f es **diferenciable Gateaux** en x_0 . Como en dimensión finita, si existe la diferencial de Fréchet, también existe la diferencial de Gateaux $d(x_0, \cdot)$ y se tiene $df(x_0, u) = Df(x_0) \cdot u$ para todo $u \in E$. El recíproco no es cierto en general, pero se demuestra que si existe la diferencial de Gateaux $d(x, \cdot)$ en todo x de un entorno U de x_0 y la aplicación $x \mapsto d(x, \cdot)$ de $U \subseteq D \subseteq E$ en $\mathcal{L}(E, F)$ es continua, entonces f es diferenciable Fréchet en x_0 y $df(x_0, u) = Df(x_0) \cdot u$ para todo $u \in E$. Dado que el cálculo de diferenciales de Gateaux es mucho más fácil, este hecho proporciona un método práctico muy útil para calcular diferenciales de Fréchet. La regla de la cadena se verifica para las diferenciales de Fréchet, como hemos apuntado antes, pero no, en general, para las de Gateaux. Esta es la razón fundamental para basar el cálculo diferencial en espacios normados sobre la diferenciabilidad más fuerte en el sentido de Fréchet.

8.5 Problemas

1. Dada una matriz A $n \times n$ se considera la función $f: \mathbb{R}^n \rightarrow \mathbb{R}$ definida por $f(x) = x^T A x$. Demostrar que

$$f(x_0 + h) - f(x_0) = h^T A x_0 + x_0^T A h + h^T A h$$

y deducir de ello que $Df(x_0) = 2x_0^T A$.

2. Probar que la cicloide es la curva que traza un punto P de una circunferencia de radio a cuando la circunferencia rueda sin deslizarse a lo largo del eje Ox . (Indicación: Figura 8.1; utilícese como parámetro t el ángulo que da la rotación de la circunferencia determinado por la vertical y el radio que, en cada posición, une el centro de la misma con el punto P .)

3. Dada la hélice $\phi(t) = (2 \cos t, 2 \sin t, t)$ se pide:

- Determinar su vector tangente.
- Comprobar que no se cumple con igualdad el teorema del valor medio en el intervalo $[0, 2\pi]$ (véase el comentario que sigue al enunciado del teorema 8.4) y que, en consecuencia, no es cierto que la velocidad media desde $t = 0$ a $t = 2\pi$ sea igual a la velocidad instantánea en un tiempo intermedio (a diferencia de lo que ocurre en una dimensión).

4. El plano normal a una curva $\phi(t)$ del espacio \mathbb{R}^3 en el punto $\phi(t_0)$ es el plano que pasa por ese punto y tiene por vector normal $\phi'(t_0)$. Determinar el plano normal a la hélice $(a \cos t, a \sin t, ht)$ en el punto $\phi(0)$.

5. Probar que si una partícula recorre una curva $\phi(t)$ manteniéndose a una distancia fija del origen, es decir, $|\phi(t)| = r$ es constante para todo t , entonces el vector posición $\phi(t)$ es perpendicular al vector velocidad $\phi'(t)$ a lo largo del movimiento. (Indicación: dérvese $|\phi(t)|^2 = \langle \phi(t), \phi(t) \rangle$.)

6. Demostrar que la velocidad (escalar) es constante si y sólo si el vector aceleración es perpendicular al vector velocidad.

7. Un sistema gradiente en un abierto $\Omega \subseteq \mathbb{R}^n$ es un sistema de la forma $x' = -\nabla F(x)$ donde $F: \Omega \rightarrow \mathbb{R}$ es de clase C^2 (es tradicional la escritura con el signo $(-)$ con el objeto de simplificar ciertas expresiones; no hay pérdida de generalidad pues, desde luego, el campo vectorial $f(x)$ es el gradiente de cierta función potencial: $f(x) = \nabla(-F(x))$). Demostrar los siguientes resultados:

- (i) Los puntos de equilibrio del sistema son los puntos críticos de la función F .
- (ii) Dado $x \in \Omega$, sea $\phi(t, x)$ la solución del sistema tal que $\phi(0, x) = x$. Si

$$\dot{F}(x) \stackrel{\text{def}}{=} \left. \frac{d}{dt} F(\phi(t, x)) \right|_{t=0}$$

entonces $\dot{F}(x) \leq 0$ para todo $x \in \Omega$ y $\dot{F}(x) = 0$ si y sólo si x es punto de equilibrio.

- (iii) En los puntos regulares (los que no son de equilibrio), las trayectorias cortan ortogonalmente a las superficies de nivel (curvas si $n = 2$) de la función F . Dibujar las trayectorias del sistema gradiente definido por $F(x, y) = x^2 + y^2$.
- (iv) Un sistema gradiente no posee trayectorias que sean curvas cerradas simples (que corresponderían a soluciones periódicas del sistema).
8. Si un campo de fuerzas $f(x)$ es un campo gradiente, o sea, tal que $f(x) = -\nabla F(x)$ para cierta $F: \Omega \rightarrow \mathbb{R}$, se dice que es *conservativo*, por la razón que enseguida veremos. La función F se dice que es la *función energía potencial* asociada al campo (una energía potencial, más bien, pues $F(x) + C$ para cualquier constante C proporciona el mismo campo de fuerzas $f(x)$). Si $\phi(t)$ es el movimiento en dicho campo de fuerzas de una partícula de masa m , la cantidad

$$\frac{1}{2} m |\phi'(t)|^2$$

es su *energía cinética* en el instante t y

$$E(t) = \frac{1}{2} m |\phi'(t)|^2 + F(\phi(t))$$

su *energía total*.

- (i) Demostrar que en un campo conservativo la energía total de cualquier movimiento es constante a lo largo del tiempo (la energía total se *conserva*. **Indicación:** Derívase $E(t)$ utilizando la regla de la cadena.)
- (ii) Comprobar que $F(x) = -K/|x|$ es una energía potencial del campo gravitatorio (8.57).

- (iii) En el caso $n = 1$, todo campo $f(x)$ es conservativo pues basta tomar como $F(x)$ una primitiva de $f(x)$ con signo cambiado (se supone que f es de clase al menos C^1 en un intervalo I de \mathbb{R}). Interpretese $F(x)$ como el *trabajo* que hay que realizar *contra* las fuerzas del campo para llevar la partícula desde una posición fija de referencia $\bar{x} \in I$ hasta la posición x . Se puede dar la misma interpretación en $n = 2$ y $n = 3$ utilizando integrales curvilíneas)

El resultado de (i) significa que para la trayectoria del sistema

$$\begin{cases} x' = y \\ y' = f(x) \end{cases}$$

que verifique $x(0) = x_0$, $y(0) = y_0$ (o sea, la correspondiente a la posición inicial x_0 y la velocidad inicial y_0), se tiene

$$\frac{1}{2}my^2 + F(x) = \frac{1}{2}my_0^2 + F(x_0)$$

relación que define *implícitamente* a la trayectoria en el plano. Apoyándose en esto, resolver la ecuación diferencial de segundo orden

$$x'' = -x$$

Capítulo 9

El teorema de la función implícita

Este capítulo está dedicado a la demostración pendiente del teorema de la función implícita y los resultados conexos relativos a la existencia de función inversa, la dependencia funcional y el método de los multiplicadores de Lagrange y de Kuhn-Tucker. El camino elegido para la demostración de aquél se basa en el potente *método de aproximaciones sucesivas* que se expone en el apartado 9.1, el cual se aplica en otras situaciones del análisis matemático y en contextos mucho más generales que el de este curso.

9.1 Aproximaciones sucesivas

Como decíamos en el apartado 4.2, el problema de si el sistema de m ecuaciones en $n + m$ variables

$$\begin{cases} F_1(x_1, \dots, x_n, y_1, \dots, y_m) = 0 \\ \vdots \\ F_m(x_1, \dots, x_n, y_1, \dots, y_m) = 0 \end{cases} \quad (9.1)$$

se puede resolver en las m variables y_1, \dots, y_m , o sea, si define o no implícitamente a las variables y_1, \dots, y_m como funciones de las variables x_1, \dots, x_n :

$$\begin{cases} y_1 = f_1(x_1, \dots, x_n) \\ \vdots \\ y_m = f_m(x_1, \dots, x_n) \end{cases} \quad (9.2)$$

se puede contemplar como el de la búsqueda de soluciones del sistema próximas a una solución conocida $(x_0, y_0) = (x_1^0, \dots, x_n^0, y_1^0, \dots, y_m^0)$.

Muchos problemas de resolución de ecuaciones (no sólo algebraicas, sino también diferenciales, integrales, etc.) se pueden plantear como **problemas de punto fijo**: dado un conjunto X y una aplicación $T: X \rightarrow X$, se dice que $\xi \in X$ es un **punto fijo** de T si $T(\xi) = \xi$. Un procedimiento de búsqueda de puntos fijos es el llamado **método de aproximaciones sucesivas**. Consiste en partir de un "valor de prueba" x_0 y calcular su transformado $x_1 = T(x_0)$. Si $x_1 = x_0$, ya hemos llegado a un punto fijo, y si $x_1 \neq x_0$ cabe esperar (bajo ciertas circunstancias) que x_1 sea mejor "valor de prueba" que x_0 , procediendo entonces a reiterar el proceso, es decir, a calcular $x_2 = T(x_1)$ y así sucesivamente:

$$x_{k+1} = T(x_k)$$

con la esperanza de que el método sea **convergente**, es decir, que la sucesión de iterantes o aproximaciones sucesivas x_k converja hacia el punto fijo. Naturalmente, el mero hecho de hablar de "convergencia" implica que en el conjunto base X ha de tener sentido hablar de cuestiones de convergencia, o sea, que X debe estar dotado de una **topología**. En el contexto presente, $X = \mathbb{R}^n$, para cierto n , y el proyecto descrito tiene perfecto sentido.

Entre las diversas hipótesis que garantizan la convergencia del método de aproximaciones sucesivas, la más conocida es la de **contractividad**, como se demuestra en el teorema que sigue, uno de los más célebres del análisis. Recordemos la siguiente

Definición 9.1 Sea X un subconjunto de \mathbb{R}^n . Una aplicación $T: X \rightarrow \mathbb{R}^n$ se dice que es **contractiva** (o que es una **contracción**) si existe una constante $\alpha < 1$ tal que

$$|T(x) - T(y)| \leq \alpha |x - y|$$

para todo par de puntos $x, y \in X$.

(Obsérvese que una aplicación contractiva es, en particular, lipschitziana y, por tanto, uniformemente continua.)

Teorema 9.1 (teorema de la aplicación contractiva o del punto fijo de Banach). Sea X un subconjunto cerrado de \mathbb{R}^n y sea $T: X \rightarrow X$ una aplicación contractiva. Entonces T tiene un único punto fijo ξ en X y, para todo $x_0 \in X$, la sucesión de iterantes $x_1 = T(x_0), \dots, x_k = T(x_{k-1})$ converge a ξ y se satisface la siguiente estimación del error

$$|x_k - \xi| \leq \frac{\alpha^k}{1 - \alpha} |x_1 - x_0| \quad (9.3)$$

Demostración. Dado $x_0 \in X$, definimos $x_k = T(x_{k-1})$. Para todo k :

$$|x_{k+1} - x_k| = |T(x_k) - T(x_{k-1})| \leq \alpha |x_k - x_{k-1}| \leq \dots \leq \alpha^k |x_1 - x_0|$$

de donde, si $m > k$, se tiene

$$\begin{aligned} |x_m - x_k| &\leq |x_m - x_{m-1}| + \dots + |x_{k+1} - x_k| \leq (\alpha^{m-1} + \dots + \alpha^k) |x_1 - x_0| = \\ &= \frac{\alpha^k - \alpha^m}{1 - \alpha} |x_1 - x_0| \leq \frac{\alpha^k}{1 - \alpha} |x_1 - x_0| \end{aligned}$$

Al ser $\alpha < 1$, se tiene que $\alpha^k \rightarrow 0$ cuando $k \rightarrow \infty$. Por tanto, para cada $\varepsilon > 0$ existe k_0 tal que

$$\frac{\alpha^k}{1 - \alpha} |x_1 - x_0| < \varepsilon$$

para todo $k \geq k_0$, y entonces

$$|x_m - x_k| \leq \frac{\alpha^k}{1 - \alpha} |x_1 - x_0| < \varepsilon$$

para todo $m, k \geq k_0$. Esto significa que $\{x_k\}$ es una sucesión de Cauchy en X que, por ser éste cerrado, converge a cierto $\xi \in X$. Se tiene:

$$\begin{aligned} |T(\xi) - \xi| &= |T(\xi) - T(x_k) + T(x_k) - x_k + x_k - \xi| \leq \\ &\leq |T(\xi) - T(x_k)| + |T(x_k) - x_k| + |x_k - \xi| \leq \\ &\leq (1 + \alpha) |x_k - \xi| + \alpha^k |x_1 - x_0| \rightarrow 0 \text{ cuando } k \rightarrow \infty \end{aligned}$$

y, por tanto, $|T(\xi) - \xi| = 0$, es decir, $T(\xi) = \xi$.

La unicidad del punto fijo es evidente, pues si ξ, η fuesen dos puntos fijos distintos tendríamos $|\xi - \eta| = |T(\xi) - T(\eta)| \leq \alpha |\xi - \eta|$, de donde $(1 - \alpha) |\xi - \eta| \leq 0$, lo que es imposible por ser $\alpha < 1$ y $|\xi - \eta| > 0$.

Observaciones:

1. Algunos de los métodos que se utilizan para la determinación de soluciones de una ecuación $f(x) = 0$, $f: \mathbb{R} \rightarrow \mathbb{R}$, pueden formularse en el marco de la teoría de puntos fijos. Por ejemplo, el llamado *método de la cuerda* consiste en lo siguiente: se parte de una aproximación inicial x_0 a la solución que se busca y trazamos la recta con pendiente m fijada que pasa por el punto $(x_0, f(x_0))$ (Fig. 9.1); esta recta corta al eje de abscisas en el punto x_1 , que, esperamos, es una mejor aproximación que x_0 a la solución buscada; se repite la operación a partir de x_1 , y así sucesivamente:

$$m = \frac{f(x_0)}{x_0 - x_1} \quad (9.4)$$

y, en general,

$$m = \frac{f(x_k)}{x_k - x_{k+1}}, \quad k = 0, 1, 2, \dots \quad (9.5)$$

o sea,

$$m(x_{k+1} - x_k) = -f(x_k), \quad k = 0, 1, 2, \dots \quad (9.6)$$

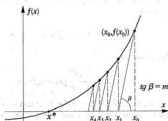


Figura 9.1

Reinterpretemos el proceso anterior como un problema de punto fijo: de (9.6) se tiene

$$x_{k+1} = x_k - \frac{f(x_k)}{m} \quad (9.7)$$

Introduciendo la aplicación

$$T(x) = x - \frac{f(x)}{m} \quad (9.8)$$

vemos que los puntos fijos de T son las soluciones de $f(x) = 0$, pues

$$x^* = T(x^*) = x^* - \frac{f(x^*)}{m}$$

es obviamente equivalente a $f(x^*) = 0$. Asimismo, se observa que (9.7) no es otra cosa que la sucesión de iteradas $x_{k+1} = T(x_k)$. Si T transforma un cierto intervalo cerrado I en sí mismo (es decir, $T(I) \subset I$) y verifica $|T'(\xi)| \leq \alpha < 1$ (o sea, $|1 - f'(\xi)/m| \leq \alpha$) para todo $\xi \in I$, entonces T es contractiva —por el teorema del valor medio— y, por el teorema de Banach, la sucesión $\{x_k\}$ converge a cierto límite $x^* \in I$ que es un punto fijo (el único en I) de T .

Si se trata de una ecuación vectorial $f(x) = 0$, $f: \mathbb{R}^n \rightarrow \mathbb{R}^n$, la generalización del esquema iterativo (9.6) sería

$$M(x_{k+1} - x_k) = -f(x_k), \quad k = 0, 1, 2, \dots \quad (9.9)$$

siendo M una matriz $n \times n$ no singular. Si se toma, en particular, $M = Df(x_0)$ ($m = f'(x_0)$ en \mathbb{R}), se tiene el llamado *método de Newton corregido*. En el *método de Newton* usual se va variando en cada caso la "pendiente" de acuerdo con el esquema

$$Df(x_k)(x_{k+1} - x_k) = -f(x_k), \quad k = 0, 1, 2, \dots \quad (9.10)$$

Se aprecia en la figura 9.2 que el método de Newton, con vistas a las aplicaciones prácticas, proporciona una convergencia más rápida, pero, como contrapartida, tiene el inconveniente de que puede ser complicada, si $n > 1$, la evaluación en cada paso de la matriz jacobiana y la resolución de (9.10).

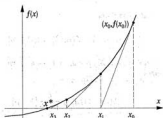


Figura 9.2

2. Obsérvese que la demostración del Teorema 9.1 vale para cualquier norma de \mathbb{R}^n . La condición es que la aplicación T sea contractiva para la norma considerada (una aplicación puede ser contractiva respecto a una norma y no serlo respecto a otra).

3. El método de aproximaciones sucesivas (a pesar de que su nombre así parece indicarlo) no es la única forma de aproximarse a un punto fijo: el método de bipartición de las demostraciones tradicionales del teorema de ceros de Bolzano da una idea de otras posibilidades. Existen otros métodos basados en el cálculo diferencial, como el de Newton (9.10) antes mencionado y otros.

En muchas aplicaciones de interés (en particular, en la demostración del teorema de la función implícita que veremos a continuación), la aplicación T

depende de uno o varios parámetros, y es importante saber en qué sentido el punto fijo cuya existencia se demuestra en el teorema 1 depende de ellos.

Teorema 9.2 (de la contracción uniforme). Sean Λ un subconjunto de \mathbb{R}^p y X un subconjunto cerrado de \mathbb{R}^n y sea $T: \Lambda \times X \rightarrow X$ una aplicación continua y uniformemente contractiva, es decir, tal que existe $\alpha < 1$ con la propiedad

$$|T(\lambda, x) - T(\lambda, y)| \leq \alpha |x - y| \quad (9.11)$$

para todo $x, y \in X$ y todo $\lambda \in \Lambda$.

1. Sea $\xi(\lambda)$ el (único) punto fijo de la aplicación contractiva $x \mapsto T(\lambda, x)$. Entonces la aplicación $\lambda \mapsto \xi(\lambda)$ de Λ en X es continua.
2. Si además existe L tal que

$$|T(\lambda, x) - T(\mu, x)| \leq L |\lambda - \mu|$$

para todo $\lambda, \mu \in \Lambda$, $x \in X$, o sea, si T es uniformemente lipschitziana en λ , entonces

$$|\xi(\lambda) - \xi(\mu)| \leq \frac{L}{1-\alpha} |\lambda - \mu| \quad (9.12)$$

es decir, la aplicación $\lambda \mapsto \xi(\lambda)$ es lipschitziana.

Demostración.

$$\begin{aligned} |\xi(\lambda) - \xi(\mu)| &= |T(\lambda, \xi(\lambda)) - T(\mu, \xi(\mu))| \leq \\ &\leq |T(\lambda, \xi(\lambda)) - T(\lambda, \xi(\mu))| + |T(\lambda, \xi(\mu)) - T(\mu, \xi(\mu))| \leq \\ &\leq \alpha |\xi(\lambda) - \xi(\mu)| + |T(\lambda, \xi(\mu)) - T(\mu, \xi(\mu))| \end{aligned}$$

de donde

$$|\xi(\lambda) - \xi(\mu)| \leq \frac{1}{1-\alpha} |T(\lambda, \xi(\mu)) - T(\mu, \xi(\mu))|$$

que tiende a 0 cuando $\lambda \rightarrow \mu$ por la continuidad de T . La segunda afirmación del teorema es evidente.

9.2 Teoremas de la función implícita e inversa

Para establecer el teorema de la función implícita para sistemas de varias ecuaciones en varias variables hay que definir las *diferenciales parciales* de una función vectorial. Sea $F: \Omega \rightarrow \mathbb{R}^m$ una función definida en un abierto Ω de \mathbb{R}^N . Contemplemos \mathbb{R}^N como el producto cartesiano $\mathbb{R}^p \times \mathbb{R}^q$ separando las $N = p + q$ variables que describen los puntos de \mathbb{R}^N en dos grupos de, respectivamente, p y q variables: $x = (x_1, \dots, x_p)$, $y = (y_1, \dots, y_q)$. Fijadas unas normas en \mathbb{R}^p y \mathbb{R}^q , que representamos simplemente por $|x|$ y $|y|$, utilizaremos en $\mathbb{R}^N \approx \mathbb{R}^p \times \mathbb{R}^q$ la norma $|(x, y)| = |x| + |y|$ (es claramente una norma en $\mathbb{R}^p \times \mathbb{R}^q$ y recordemos que en cuestiones de continuidad y diferenciabilidad podemos usar en estos espacios euclídeos cualquier norma pues todas ellas son equivalentes).

Dado $(x_0, y_0) \in \Omega$, $x_0 \in \mathbb{R}^p$, $y_0 \in \mathbb{R}^q$, se tienen definidas las aplicaciones parciales

$$x \mapsto F(x, y_0), \quad y \mapsto F(x_0, y)$$

Pues bien, si la función parcial $x \mapsto F(x, y_0)$ es diferenciable en x_0 , su diferencial, elemento de $\mathcal{L}(\mathbb{R}^p, \mathbb{R}^m)$ que denotamos por $D_x F(x_0, y_0)$, se dice que es la *diferencial parcial de F respecto a x en (x_0, y_0)* . Análogamente, la *diferencial parcial de F respecto a y en (x_0, y_0)* es la diferencial $D_y F(x_0, y_0) \in \mathcal{L}(\mathbb{R}^q, \mathbb{R}^m)$ de la aplicación parcial $y \mapsto F(x_0, y)$ en el punto y_0 .

Proposición 9.1 Si $F: \mathbb{R}^p \times \mathbb{R}^q \supseteq \Omega \rightarrow \mathbb{R}^m$ es diferenciable en (x_0, y_0) , entonces existen las diferenciales parciales $D_x F(x_0, y_0)$, $D_y F(x_0, y_0)$ y se tiene

$$DF(x_0, y_0) \cdot (h, k) = D_x F(x_0, y_0) \cdot h + D_y F(x_0, y_0) \cdot k \quad (9.13)$$

para todo par $h = (h_1, \dots, h_p) \in \mathbb{R}^p$, $k = (k_1, \dots, k_q) \in \mathbb{R}^q$.

Demostración: Sea A la aplicación lineal $\mathbb{R}^p \rightarrow \mathbb{R}^m$ definida por

$$A(h) = D_x F(x_0, y_0) \cdot (h, 0)$$

Por la diferenciabilidad de F en (x_0, y_0) se tiene:

$$\begin{aligned} \lim_{h \rightarrow 0} \frac{1}{|h|} [F(x_0 + h, y_0) - F(x_0, y_0) - A(h)] &= \\ &= \lim_{(h,0) \rightarrow (0,0)} \frac{1}{|(h,0)|} [F(x_0 + h, y_0) - F(x_0, y_0) - D_x F(x_0, y_0) \cdot (h, 0)] = 0 \end{aligned}$$

y, en consecuencia, $x \mapsto F(x, y_0)$ es diferenciable en x_0 y su diferencial es $h \mapsto DF(x_0, y_0) \cdot (h, 0)$. De la misma manera se prueba la diferenciabilidad parcial respecto a y y la fórmula (9.13) deriva de la linealidad de $DF(x_0, y_0)$:

$$\begin{aligned} DF(x_0, y_0) \cdot (h, k) &= DF(x_0, y_0) \cdot (h, 0) + DF(x_0, y_0) \cdot (0, k) = \\ &= D_x F(x_0, y_0) \cdot h + D_y F(x_0, y_0) \cdot k \end{aligned}$$

$DF(x_0, y_0)$ está representada por la matriz jacobiana $m \times (p+q)$

$$\begin{pmatrix} \frac{\partial F_1}{\partial x_1} & \cdots & \frac{\partial F_1}{\partial x_p} & \frac{\partial F_1}{\partial y_1} & \cdots & \frac{\partial F_1}{\partial y_q} \\ \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\ \frac{\partial F_m}{\partial x_1} & \cdots & \frac{\partial F_m}{\partial x_p} & \frac{\partial F_m}{\partial y_1} & \cdots & \frac{\partial F_m}{\partial y_q} \end{pmatrix}$$

evaluada en (x_0, y_0) . Haciéndola actuar sobre vectores $(h, 0)$ y $(0, k)$ vemos que las representaciones matriciales de $D_x F(x_0, y_0)$ y $D_y F(x_0, y_0)$ son, respectivamente, las submatrices de la anterior

$$\begin{pmatrix} \frac{\partial F_1}{\partial x_1} & \cdots & \frac{\partial F_1}{\partial x_p} \\ \vdots & \cdots & \vdots \\ \frac{\partial F_m}{\partial x_1} & \cdots & \frac{\partial F_m}{\partial x_p} \end{pmatrix} \text{ y } \begin{pmatrix} \frac{\partial F_1}{\partial y_1} & \cdots & \frac{\partial F_1}{\partial y_q} \\ \vdots & \cdots & \vdots \\ \frac{\partial F_m}{\partial y_1} & \cdots & \frac{\partial F_m}{\partial y_q} \end{pmatrix}$$

En realidad, la proposición anterior es la generalización del resultado del capítulo 3 que establece que una función diferenciable posee derivadas parciales. En efecto, lo que se ha expuesto para una descomposición de \mathbb{R}^N en producto cartesiano de dos factores se extiende sin dificultad a cualquier número de ellos: $\mathbb{R}^N = \mathbb{R}^{p_1} \times \cdots \times \mathbb{R}^{p_r}$, $p_1 + \cdots + p_r = N$; en particular, a N factores: $\mathbb{R}^N = \mathbb{R} \times \cdots \times \mathbb{R}$ y como en dimensión 1 las nociones de diferenciabilidad y derivabilidad coinciden, obtenemos como caso particular de lo anterior el resultado del capítulo 3 mencionado (así como el correspondiente a funciones vectoriales del capítulo 8). Y esta observación nos muestra también que el recíproco de la proposición 9.1 no es cierto, o sea, pueden existir las diferenciales parciales en (x_0, y_0) y sin embargo no ser diferenciable en (x_0, y_0) la función F : basta recordar los ejemplos que conocemos de funciones de dos variables que tienen derivadas parciales en un punto y que no son diferenciables en él. Si se tiene, como cabe esperar, la generalización de los teoremas 3.2 y 8.1:

Proposición 9.2 *F es de clase C^1 en Ω si y sólo si existen las diferenciales parciales $D_x F(x, y)$, $D_y F(x, y)$ en todo $(x, y) \in \Omega$ y son continuas en Ω las*

aplicaciones

$$(x, y) \mapsto D_x F(x, y) \in \mathcal{L}(\mathbb{R}^p, \mathbb{R}^m), \quad (x, y) \mapsto D_y F(x, y) \in \mathcal{L}(\mathbb{R}^q, \mathbb{R}^m)$$

Demostración: La necesidad es clara teniendo en cuenta las representaciones matriciales: si F es de clase C^1 —o sea, si todas sus derivadas parciales son continuas en Ω — entonces es diferenciable, con lo que, por la proposición 9.1, existen las diferenciales parciales, las derivadas parciales de las funciones parciales $x \mapsto F(x, y)$, $y \mapsto F(x, y)$ son continuas en Ω y, por los argumentos expuestos al final del apartado 8.1, las aplicaciones

$$\begin{aligned}(x, y) &\mapsto DF(x, y) \in \mathcal{L}(\mathbb{R}^{p+q}, \mathbb{R}^m) \\(x, y) &\mapsto D_x F(x, y) \in \mathcal{L}(\mathbb{R}^p, \mathbb{R}^m) \\(x, y) &\mapsto D_y F(x, y) \in \mathcal{L}(\mathbb{R}^q, \mathbb{R}^m)\end{aligned}$$

son continuas en Ω (naturalmente, también se puede argumentar de una manera más directa y abstracta, sin pasar por las derivadas parciales, concluyendo de la continuidad de la primera de estas aplicaciones la continuidad de las otras dos teniendo en cuenta la expresión de las diferenciales parciales en términos de la diferencial de F que se estableció en la proposición 9.1).

En cuanto a la suficiencia, se propone, lógicamente, como diferencial de F en (x_0, y_0) la aplicación lineal

$$(h, k) \mapsto D_x F(x_0, y_0) \cdot h + D_y F(x_0, y_0) \cdot k$$

y se escribe, como en la demostración del teorema 3.2,

$$\begin{aligned}&|F(x_0 + h, y_0 + k) - F(x_0, y_0) - [D_x F(x_0, y_0) \cdot h + D_y F(x_0, y_0) \cdot k]| \leq \\&\leq |F(x_0 + h, y_0) - F(x_0, y_0) - D_x F(x_0, y_0) \cdot h| + \\&+ |F(x_0 + h, y_0 + k) - F(x_0 + h, y_0) - D_y F(x_0, y_0) \cdot k| \leq \text{(como en el} \\&\text{corolario 8.1)} \leq \sup_{0 < \theta < 1} \|D_x F(x_0 + \theta h, y_0) - D_x F(x_0, y_0)\| |h| + \\&+ \sup_{0 < \theta < 1} \|D_y F(x_0 + h, y_0 + \theta k) - D_y F(x_0, y_0)\| |k|\end{aligned}$$

Por la continuidad de las aplicaciones $(x, y) \mapsto D_x F(x, y)$, $(x, y) \mapsto D_y F(x, y)$, dado $\varepsilon > 0$, existe $\delta > 0$ tal que, si $|x - x_0| < \delta$, $|y - y_0| < \delta$, entonces

$$\|D_x F(x, y) - D_x F(x_0, y_0)\| < \varepsilon, \quad \|D_y F(x, y) - D_y F(x_0, y_0)\| < \varepsilon$$

lo que implica, para $|h| < \delta$, $|k| < \delta$,

$$|F(x_0 + h, y_0 + k) - F(x_0, y_0) - [D_x F(x_0, y_0) \cdot h + D_y F(x_0, y_0) \cdot k]| < \varepsilon(|h| + |k|)$$

y, de aquí, que F es diferenciable en (x_0, y_0) y que su diferencial es efectivamente la aplicación lineal $(h, k) \mapsto D_x F(x_0, y_0) \cdot h + D_y F(x_0, y_0) \cdot k$.

También nos va a ser útil la adaptación de la proposición 8.5 a las diferenciales parciales:

Proposición 9.3 Sea $F : \mathbb{R}^p \times \mathbb{R}^q \supseteq \Omega \rightarrow \mathbb{R}^m$ tal que la diferencial parcial $D_y F(x, y)$ existe en todo $(x, y) \in \Omega$ y la aplicación $(x, y) \mapsto D_y F(x, y)$ es continua de Ω en $\mathcal{L}(\mathbb{R}^q, \mathbb{R}^m)$. Entonces, para todo $(x_0, y_0) \in \Omega$ y todo $\gamma > 0$ existen $\delta > 0, \varepsilon > 0$ tales que $B_\delta(x_0) \times B_\varepsilon(y_0) \subset \Omega$,

$$|F(x, y_2) - F(x, y_1) - D_y F(x_0, y_0) \cdot (y_2 - y_1)| \leq \gamma |y_2 - y_1|$$

para todo $x \in B_\delta(x_0)$ y todo par $y_1, y_2 \in B_\varepsilon(y_0)$, y, como consecuencia,

$$|F(x, y_2) - F(x, y_1)| \leq (\|D_y F(x_0, y_0)\| + \gamma) |y_2 - y_1|$$

para $(x, y_1), (x, y_2) \in B_\delta(x_0) \times B_\varepsilon(y_0)$.

En efecto, se considera, como en la demostración de la proposición 8.5, la aplicación $G(x, y) = F(x, y) - D_y F(x_0, y_0) \cdot y$. Así, se tiene $D_y G(x_0, y_0) = D_y F(x_0, y_0) - D_y F(x_0, y_0) = 0$ (la matriz cero $m \times q$) y se argumenta como allí a partir de la continuidad de la aplicación $(x, y) \mapsto D_y G(x, y)$.

Con estos preparativos, enunciamos ya el teorema de la función implícita:

Teorema 9.3 Sea $F : \Omega \rightarrow \mathbb{R}^m$ una función de clase C^1 en un abierto Ω de $\mathbb{R}^n \times \mathbb{R}^m$. Supongamos que el punto $(x_0, y_0) \in \Omega$ es tal que:

1. $F(x_0, y_0) = 0$.
2. $D_y F(x_0, y_0) \in \mathcal{L}(\mathbb{R}^m)$ es invertible.

Entonces existen $\delta > 0, \varepsilon > 0$ y una función $f : B_\delta(x_0) \rightarrow B_\varepsilon(y_0)$, de clase C^1 tales que

1. $f(x_0) = y_0$.
2. $F(x, f(x)) = 0$ para todo $x \in B_\delta(x_0)$.

3. Para cada $x \in B_\delta(x_0)$, $y = f(x)$ es la única solución de la ecuación $F(x, y) = 0$ que pertenece a $B_\varepsilon(y_0)$.
4. La diferencial de f en $x \in B_\delta(x_0)$ está dada por

$$Df(x) = -(D_y F(x, f(x)))^{-1} D_x F(x, f(x)), \quad (9.14)$$

Así pues, y en términos prácticos, el teorema establece que si se tiene un sistema de m ecuaciones y $n + m$ variables:

$$F_1(x_1, \dots, x_n, y_1, \dots, y_m) = 0$$

$$\vdots$$

$$F_m(x_1, \dots, x_n, y_1, \dots, y_m) = 0$$

y es distinto de cero el determinante jacobiano de $F = (F_1, \dots, F_m)$ respecto a las variables (y_1, \dots, y_m) :

$$\det \left(\frac{\partial F_i}{\partial y_j} \right), \quad i, j = 1, \dots, m$$

evaluado en $(x_0, y_0) = (x_0^0, \dots, x_n^0, y_1^0, \dots, y_m^0)$, donde (x_0, y_0) es una solución conocida del sistema, entonces es posible, al menos teóricamente, expresar las variables y_1, \dots, y_m en función de las variables x_1, \dots, x_n en un entorno de (x_0, y_0) .

Demostración. Según adelantábamos al comienzo del capítulo, la idea de la demostración es plantear el problema de la existencia de funciones $y = f(x)$ definidas implícitamente por la ecuación $F(x, y) = 0$ como el de la búsqueda de soluciones de la ecuación próximas a la solución conocida (x_0, y_0) . Para ello, vamos a aplicar el método de Newton corregido (9.9) a la ecuación $F(x, y) = 0$, con $(x, y) \in \Omega \subset \mathbb{R}^n \times \mathbb{R}^m$, $F: \Omega \rightarrow \mathbb{R}^m$, considerando x como un parámetro, con la idea de aplicar el teorema de la contracción uniforme. Para simplificar (no se pierde generalidad con ello) hagamos $x_0 = 0$, $y_0 = 0$, llamemos $M = D_y F(0, 0)$ y planteemos el esquema iterativo

$$M(y_{k+1} - y_k) = -F(x, y_k), \quad k = 0, 1, 2, \dots \quad (9.15)$$

Si se demuestra que $\{y_k\}$ converge a un límite $f(x)$, entonces, por la continuidad de F , se tendrá $F(x, f(x)) = 0$. Escribiendo (9.15) en la forma

$$y_{k+1} = y_k - M^{-1}F(x, y_k) = M^{-1}[My_k - F(x, y_k)] \quad (9.16)$$

veamos que $f(x)$ puede interpretarse como punto fijo de la aplicación

$$T(x, y) = M^{-1}[My - F(x, y)] \quad (9.17)$$

en la que x juega el papel de parámetro (pues

$$T(x, \xi) = \xi \Leftrightarrow \xi = M^{-1}[M\xi - F(x, \xi)] = \xi - M^{-1}F(x, \xi) \Leftrightarrow F(x, \xi) = 0$$

por ser M^{-1} no singular). Los puntos fijos de T son, pues, los ceros de $F(x, y)$, y nuestro objetivo es, como se apuntaba, intentar aplicar a T el teorema de la contracción uniforme. Para ello, probaremos que, para todo $\varepsilon > 0$ suficientemente pequeño, se puede encontrar $\delta > 0$ tal que T envía la bola $\bar{B}_\varepsilon(0)$ en $B_\varepsilon(0) \subset \bar{B}_\varepsilon(0)$ para todo $x \in B_\delta(0)$ y que T es contractiva en $\bar{B}_\varepsilon(0)$ uniformemente respecto a $x \in B_\delta(0)$. Veamos: puesto que $(x, y) \mapsto D_y F(x, y)$ es continua por hipótesis (proposición 9.2) se tiene por la proposición 9.3 que, dado $\gamma > 0$, existen $\delta > 0, \varepsilon_1 > 0$ de modo que $B_\delta(0) \times B_{\varepsilon_1}(0) \subset \Omega$ y

$$\begin{aligned} |T(x, y_2) - T(x, y_1)| &= \\ &= \left| (D_y F(0, 0))^{-1} [D_y F(0, 0)(y_2 - y_1) - (F(x, y_2) - F(x, y_1))] \right| \leq \\ &\leq \left\| (D_y F(0, 0))^{-1} \right\| |D_y F(0, 0)(y_2 - y_1) - (F(x, y_2) - F(x, y_1))| \leq \\ &\leq \gamma \left\| (D_y F(0, 0))^{-1} \right\| |y_2 - y_1| \end{aligned}$$

para todo par $(x, y_1), (x, y_2) \in B_\delta(0) \times B_{\varepsilon_1}(0)$. Elijiendo $\gamma = \alpha \|M^{-1}\|^{-1}$ para un número $\alpha \in (0, 1)$ fijado resulta

$$|T(x, y_2) - T(x, y_1)| \leq \alpha |y_2 - y_1| \quad (9.18)$$

para todo $x \in B_\delta(0)$ y todo par $y_1, y_2 \in B_{\varepsilon_1}(0)$, y, desde luego, fijado un $\varepsilon < \varepsilon_1$, para todo $x \in B_\delta(0)$ y todo par $y_1, y_2 \in \bar{B}_\varepsilon(0)$ (necesitamos una bola cerrada para aplicar el teorema 9.2). Por otra parte,

$$|T(x, 0)| = \|M^{-1}F(x, 0)\| \leq (\text{pues } F(0, 0) = 0) \leq \|M^{-1}\| |F(x, 0) - F(0, 0)|$$

Como F es continua podemos conseguir, restringiendo δ si es preciso, que

$$\|M^{-1}\| |F(x, 0) - F(0, 0)| < (1 - \alpha)\varepsilon$$

de donde, junto con (9.18), se obtiene, para $x \in B_\delta(0), y \in \bar{B}_\varepsilon(0)$,

$$\begin{aligned} |T(x, y)| &\leq |T(x, y) - T(x, 0)| + |T(x, 0)| < \\ &< \alpha |y| + (1 - \alpha)\varepsilon \leq \alpha\varepsilon + (1 - \alpha)\varepsilon = \varepsilon \end{aligned}$$

es decir, según nos proponíamos, $T(\bar{B}_\varepsilon(0)) \subseteq B_\varepsilon(0) \subset \bar{B}_\varepsilon(0)$ para todo $x \in B_\delta(0)$ y, por (9.18), T es contractiva en $\bar{B}_\varepsilon(0)$ uniformemente respecto a $x \in B_\delta(0)$. Existe entonces, por el teorema de la contracción uniforme, una

función continua $f: B_\delta(0) \rightarrow \bar{B}_\varepsilon(0)$ tal que $T(x, f(x)) = f(x)$, o sea, tal que $F(x, f(x)) = 0$, y que verifica que, para cada $x \in B_\delta(0)$, $y = f(x)$ es la única solución de $F(x, y) = 0$ que pertenece a $\bar{B}_\varepsilon(0)$. Como, de hecho, $T(\bar{B}_\varepsilon(0)) \subseteq B_\varepsilon(0)$, según (9.19), se tiene $|T(x, f(x))| = |f(x)| < \varepsilon$ para todo $x \in B_\delta(0)$, es decir, f envía $B_\delta(0)$ en $B_\varepsilon(0)$ según afirma el enunciado del teorema. (Obsérvese que para esta parte de la demostración no es preciso que existan derivadas en las variables x , sino sólo que $(x, y) \mapsto D_y F(x, y)$ sea continua.)

Pasemos finalmente a la cuestión de la diferenciabilidad. Si suponemos, de momento, que f es efectivamente diferenciable en $B_\delta(0)$ y aplicamos (como ya hacíamos en el apartado 4.2) la regla de la cadena a la función

$$x \mapsto \phi(x) \stackrel{\text{def}}{=} F(x, f(x))$$

que es constante (igual a cero) en $B_\delta(0)$, se obtiene

$$D\phi(x) = D_x F(x, f(x)) + D_y F(x, f(x))Df(x) = 0 \quad (9.19)$$

para todo $x \in B_\delta(0)$. Por hipótesis, $D_y F(x_0, y_0) \in \mathcal{L}(\mathbb{R}^m)$ es invertible, es decir, $\det D_y F(x_0, y_0) \neq 0$; por otra parte, $\det D_y F(x, y)$ es una función continua por ser una suma de productos de derivadas primeras de F que, por hipótesis, son funciones continuas; en consecuencia, $\det D_y F(x, y) \neq 0$, o sea, $D_y F(x, y)$ es invertible, para todo $(x, y) \in B_\delta(0) \times B_\varepsilon(0)$ (reduciendo δ y ε si es necesario). Se tiene entonces de (9.20)

$$Df(x) = -[D_y F(x, f(x))]^{-1} D_x F(x, f(x))$$

es decir, $L \stackrel{\text{def}}{=} -[D_y F(x, f(x))]^{-1} D_x F(x, f(x))$ es la única posible candidata para diferencial de $f(x)$ en $x \in B_\delta(0)$. Para demostrar que efectivamente lo es, bastará comprobar que

$$\frac{|f(x+h) - f(x) - L(h)|}{|h|} \rightarrow 0 \text{ cuando } |h| \rightarrow 0 \quad (9.20)$$

es decir, que para todo $\rho > 0$ existe $\eta > 0$ tal que si $|h| < \eta$ se tiene

$$|f(x+h) - f(x) - L(h)| < \rho |h| \quad (9.21)$$

Teniendo en cuenta que para h pequeño se tiene $F(x+h, f(x+h)) = F(x, f(x)) = 0$, obtenemos

$$\begin{aligned} f(x+h) - f(x) - L(h) &= \\ &= f(x+h) - f(x) + D_y F(x, f(x))^{-1} D_x F(x, f(x)) \cdot h = \\ &= [D_y F(x, f(x))]^{-1} [D_x F(x, f(x)) \cdot h + D_y F(x, f(x)) \cdot (f(x+h) - f(x))] = \\ &= -[D_y F(x, f(x))]^{-1} [F(x+h, f(x+h)) - F(x, f(x)) - \\ &\quad - D_x F(x, f(x))h - D_y F(x, f(x))(f(x+h) - f(x))] \end{aligned} \quad (9.22)$$

Por otra parte, por la diferenciabilidad de $F(x, y)$ en (x, y) se tiene que (véase (9.13))

$$F(x+h, y+k) - F(x, y) - D_x(x, y) \cdot h - D_y(x, y) \cdot k = R(h, k) \quad (9.23)$$

con

$$\frac{|R(h, k)|}{|h| + |k|} \rightarrow 0 \quad \text{cuando } (h, k) \rightarrow (0, 0).$$

Dado $\mu \in (0, 1)$, sea $\sigma > 0$ tal que

$$|R(h, k)| \leq \frac{\mu}{\|D_y F(x, f(x))\|^{-1}} (|h| + |k|)$$

para todo (h, k) tal que $|h| + |k| < \sigma$. Puesto que f es continua, existe $\eta > 0$, $\eta < \sigma/2$, tal que $|f(x+h) - f(x)| < \sigma/2$ siempre que $|h| < \eta$. Entonces, si $|h| < \eta$, se tendrá $|h| + |f(x+h) - f(x)| < \sigma/2 + \sigma/2 = \sigma$, y, por (9.22),

$$\begin{aligned} |f(x+h) - f(x) - L(h)| &\leq \|D_y F(x, f(x))\|^{-1} |R(h, f(x+h) - f(x))| < \\ &\leq \mu (|h| + |f(x+h) - f(x)|) \\ &\leq \mu (|h| + |f(x+h) - f(x) - L(h)| + |L(h)|) \end{aligned}$$

de donde

$$(1 - \mu) |f(x+h) - f(x) - L(h)| \leq \mu (1 + \|L\|) |h|$$

es decir

$$|f(x+h) - f(x) - L(h)| \leq \frac{\mu(1 + \|L\|)}{1 - \mu} |h| \quad (9.24)$$

Dado, entonces, un $\rho > 0$ arbitrario, tomamos $\mu \in (0, 1)$ tal que

$$\frac{\mu(1 + \|L\|)}{1 - \mu} \leq \rho$$

y obtenemos el $\eta > 0$ correspondiente al μ elegido. Para ese valor de η se tiene (9.24) y, por tanto, (9.21).

Finalmente, la función $f(x)$ es de clase C^1 . En efecto, $[D_y F(x, f(x))]^{-1} = \text{adj}(D_y F(x, f(x))) / \det D_y F(x, f(x))$; los elementos de $\text{adj}(D_y F(x, f(x)))$ son sumas finitas de elementos de $D_y F(x, f(x))$, los cuales son funciones continuas; $\det D_y F(x, f(x))$ es, como decíamos antes, continua y distinta de cero; en consecuencia, los elementos de la matriz $[D_y F(x, f(x))]^{-1}$

son funciones continuas; lo mismo ocurre, por hipótesis, con los de $D_x F(x, f(x))$, por lo que las derivadas parciales de $f(x)$, elementos de la matriz producto

$$Df(x) = -[D_y F(x, f(x))]^{-1} D_x F(x, f(x))$$

serán también funciones continuas.

NOTA. Si F es de clase C^k , $k > 1$, entonces f también es de clase C^k . En efecto, si F es de clase C^k , sus derivadas parciales primeras son de clase C^{k-1} y, como en la argumentación anterior, eso mismo ocurrirá con los elementos de la matriz producto

$$D\varphi(x) = -[D_y F(x, f(x))]^{-1} D_x F(x, f(x))$$

EJEMPLO 1. (Véase el apartado 4.2.) El sistema de ecuaciones $F(x, y, z) = 0$ dado por

$$\begin{cases} F_1(x, y, z) \stackrel{\text{def}}{=} x + x^2 + y^3 - z = 0 \\ F_2(x, y, z) \stackrel{\text{def}}{=} xz^2 + y = 0 \end{cases}$$

define una función vectorial implícita

$$(y, z) = f(x) = (f_1(x), f_2(x))$$

de clase C^∞ en las proximidades del punto $(0, 0, 0)$ pues la matriz jacobiana

$$D_{y,z} F(x, y, z) = \begin{pmatrix} 3y^2 & -1 \\ 1 & 2xz \end{pmatrix}$$

es invertible en dicho punto. La diferencial de $f(x)$ está dada por

$$\begin{aligned} Df(x) &= -[D_{y,z} F(x, f_1(x), f_2(x))]^{-1} D_x F(x, f_1(x), f_2(x)) = \\ &= - \begin{pmatrix} 3(f_1(x))^2 & -1 \\ 1 & 2x f_2(x) \end{pmatrix}^{-1} \begin{pmatrix} 1 + 2x \\ (f_2(x))^2 \end{pmatrix} = \\ &= \frac{-1}{6x(f_1(x))^2 f_2(x) + 1} \begin{pmatrix} (2x + 4x^2 + f_2(x)) f_2(x) \\ -1 - 2x + 3(f_1(x))^2 (f_2(x))^2 \end{pmatrix} \end{aligned}$$

es decir

$$\begin{aligned} \frac{df_1}{dx} &= \frac{(2x + 4x^2 + f_2(x)) f_2(x)}{6x(f_1(x))^2 f_2(x) + 1} \\ \frac{df_2}{dx} &= \frac{1 + 2x - 3(f_1(x))^2 (f_2(x))^2}{6x(f_1(x))^2 f_2(x) + 1} \end{aligned}$$

En la práctica, para dimensiones bajas, se procede más fácilmente derivando implícitamente en las ecuaciones dadas y, sin necesidad de invertir matrices, resolviendo en las derivadas de las funciones implícitas:

$$\begin{aligned} 1 + 2x + 3y^2 \frac{dy}{dx} - \frac{dz}{dx} &= 0 \\ z^2 + 2xz \frac{dz}{dx} + \frac{dy}{dx} &= 0 \end{aligned}$$

de donde

$$\frac{dy}{dx} = -z \frac{2x + 4x^2 + z}{6xy^2z + 1}, \quad \frac{dz}{dx} = \frac{1 + 2x - 3y^2z^2}{6xy^2z + 1}$$

Así lo hicimos para este ejemplo en el apartado 4.2.

EJEMPLO 2. (Apartado 4.2) El sistema de ecuaciones $F(x, y, u, v) = 0$ dado por

$$\begin{cases} F_1(x, y, u, v) \stackrel{\text{def}}{=} xv^2 + y^2u^3 - 1 = 0 \\ F_2(x, y, z) \stackrel{\text{def}}{=} 2xy^3 + u^2v = 0 \end{cases}$$

define implícitamente una función vectorial implícita de clase C^∞ :

$$\begin{pmatrix} u \\ v \end{pmatrix} = f(x, y) = \begin{pmatrix} f_1(x, y) \\ f_2(x, y) \end{pmatrix}$$

en las proximidades del punto $(0, 1, 1, 0)$ ya que

$$\det D_{u,v}F(0, 1, 1, 0) = \begin{vmatrix} 3y^2u^2 & 2xv \\ 2uv & u^2 \end{vmatrix}_{(0,1,1,0)} = 3 \neq 0$$

La diferencial de $f(x, y)$ está dada por

$$\begin{aligned} & \begin{pmatrix} \frac{\partial f_1}{\partial x} & \frac{\partial f_1}{\partial y} \\ \frac{\partial f_2}{\partial x} & \frac{\partial f_2}{\partial y} \end{pmatrix} = Df(x, y) = \\ &= -[D_{u,v}F(x, y, f_1(x, y), f_2(x, y))]^{-1} D_{x,y}F(x, y, f_1(x, y), f_2(x, y)) = \\ &= - \begin{pmatrix} 3y^2(f_1(x, y))^2 & 2xf_2(x, y) \\ 2f_1(x, y)f_2(x, y) & (f_1(x, y))^2 \end{pmatrix}^{-1} \begin{pmatrix} (f_2(x, y))^2 & 2y(f_1(x, y))^3 \\ 2y^3 & 6xy^2 \end{pmatrix} = \\ &= \frac{1}{f_1(x, y)(-3y^2(f_1(x, y))^3 + 4x(f_2(x, y))^2)} \cdot \\ & \cdot \begin{pmatrix} f_2(x, y)(f_1^2(x, y)f_2(x, y) - 4xy^3) & -2y(-f_1^2(x, y) + 6x^2yf_2(x, y)) \\ -2f_1(x, y)(f_2^2(x, y) - 3y^5f_1(x, y)) & -2yf_1^2(x, y)(2f_1^2(x, y)f_2(x, y) - 9xy^3) \end{pmatrix} \end{aligned}$$

y de aquí se obtienen las derivadas parciales $\frac{\partial f_i}{\partial x}, \frac{\partial f_i}{\partial y}, i = 1, 2$. Como antes, estas derivadas se obtienen más fácilmente en la práctica derivando implícitamente en las ecuaciones dadas y resolviendo en las derivadas de las funciones implícitas (como ya indicamos en el apartado 4.2):

$$\begin{aligned} F_x + F_u \frac{\partial u}{\partial x} + F_v \frac{\partial v}{\partial x} &= v^2 + 3y^2 u^2 \frac{\partial u}{\partial x} + 2xv \frac{\partial v}{\partial x} = 0 \\ G_x + G_u \frac{\partial u}{\partial x} + G_v \frac{\partial v}{\partial x} &= 2y^3 + 2uv \frac{\partial u}{\partial x} + u^2 \frac{\partial v}{\partial x} = 0 \end{aligned}$$

y en consecuencia:

$$\frac{\partial u}{\partial x} = - \frac{\frac{\partial(F,G)}{\partial(x,v)}}{\frac{\partial(F,G)}{\partial(u,v)}} = - \frac{\begin{vmatrix} v^2 & 2xv \\ 2y^3 & u^2 \end{vmatrix}}{\begin{vmatrix} 3y^2 u^2 & 2xv \\ 2uv & u^2 \end{vmatrix}}, \quad \frac{\partial v}{\partial x} = - \frac{\frac{\partial(F,G)}{\partial(u,x)}}{\frac{\partial(F,G)}{\partial(u,v)}} = - \frac{\begin{vmatrix} 3y^2 u^2 & v^2 \\ 2uv & 2y^3 \end{vmatrix}}{\begin{vmatrix} 3y^2 u^2 & 2xv \\ 2uv & u^2 \end{vmatrix}}$$

De manera análoga, derivando en el sistema dado respecto a x :

$$\frac{\partial u}{\partial y} = - \frac{\frac{\partial(F,G)}{\partial(y,v)}}{\frac{\partial(F,G)}{\partial(u,v)}} = - \frac{\begin{vmatrix} 2yu^3 & 2xv \\ 6xy^2 & u^2 \end{vmatrix}}{\begin{vmatrix} 3y^2 u^2 & 2xv \\ 2uv & u^2 \end{vmatrix}}, \quad \frac{\partial v}{\partial y} = - \frac{\frac{\partial(F,G)}{\partial(u,y)}}{\frac{\partial(F,G)}{\partial(u,v)}} = - \frac{\begin{vmatrix} 3y^2 u^2 & 2yu^3 \\ 2uv & 6xy^2 \end{vmatrix}}{\begin{vmatrix} 3y^2 u^2 & 2xv \\ 2uv & u^2 \end{vmatrix}}$$

EJEMPLO 3. Un sistema lineal

$$Ax = b$$

siendo A una matriz $m \times n$ con $m < n$, define implícitamente a m variables como funciones de las $n - m$ restantes en las proximidades de cualquier solución cuando su matriz jacobiana, que está formada por las columnas de A correspondientes a las m variables dependientes, sea invertible. En consecuencia, si A es una matriz de rango completo, el sistema lineal define una función vectorial —lineal— de $n - m$ variables.

El teorema de la función inversa

Como vimos en el capítulo 4, el teorema de la función inversa se deriva inmediatamente del de la función implícita aplicado a la ecuación

$$f(x) - y = 0 :$$

Teorema 9.4 (Teorema de la función inversa) Sea $f: \Omega \rightarrow \mathbb{R}^n$ una función de clase $C^1(\Omega)$, Ω un abierto de \mathbb{R}^n . Supongamos que $Df(x_0)$ es invertible, o sea, que el jacobiano

$$\det \left(\frac{\partial f_i}{\partial x_j}(x_0) \right), \quad i, j = 1, \dots, n$$

es distinto de cero. Entonces existe un $\delta > 0$ tal que la función inversa f^{-1} está definida y es de clase C^1 en $B_\delta(f(x_0))$. La diferencial de $g(y) = f^{-1}(f(x))$ en $y = f(x) \in B_\delta(f(x_0))$ es $Dg(y) = [Df(x)]^{-1}$.

NOTAS. 1. Como en el teorema de la función implícita, si f es de clase C^k , f^{-1} también es de clase C^k .

2. Obsérvese que, bajo las hipótesis del teorema de la función inversa, el punto $y_0 = f(x_0)$ es un punto interior del conjunto $f(\Omega)$. En consecuencia, si $Df(x)$ es invertible en todo punto del abierto $\Omega \subseteq \mathbb{R}^n$, el conjunto $f(\Omega)$ es abierto.

9.3 Dependencia funcional

Se trata de demostrar aquí el resultado sobre dependencia funcional que habíamos enunciado sin demostración en el capítulo 4. Recordemos en primer lugar las definiciones:

Sean $f_1(x), \dots, f_m(x)$, $x = (x_1, \dots, x_n)$, m funciones de clase C^1 en un abierto $\Omega \subseteq \mathbb{R}^n$. Se dice que una función $f_k(x)$ depende funcionalmente en Ω de las funciones $f_i(x)$, $i = 1, \dots, m$, $i \neq k$, si existe una función φ de clase C^1 tal que $f_k(x) = \varphi(f_1(x), \dots, f_{k-1}(x), f_{k+1}(x), \dots, f_m(x))$ para todo $x \in \Omega$. Se dice que las funciones $f_1(x), \dots, f_m(x)$ son funcionalmente dependientes en Ω si una de ellas, al menos, depende funcionalmente de las restantes. En caso contrario, son funcionalmente independientes.

Ya vimos en el capítulo 4 que para que $f_1(x), \dots, f_m(x)$ sean funcionalmente dependientes en Ω es condición necesaria que el rango de la matriz jacobiana

$$Df(x) = \left(\frac{\partial f_i}{\partial x_j}(x) \right), \quad i = 1, \dots, m, \quad j = 1, \dots, n$$

sea estrictamente inferior a m en todo punto $x \in \Omega$ ($f = (f_1, \dots, f_m)$). Quedó pendiente la demostración de la siguiente condición suficiente de dependencia funcional local:

Teorema 9.5 Sean $f_i(x)$, $i = 1, \dots, m$, m funciones de clase C^1 en un entorno $U(x_0)$ del punto $x_0 \in \mathbb{R}^n$ y supongamos que

$$\text{rango} \left(\frac{\partial f_i}{\partial x_j}(x_0) \right) = r, \quad 0 < r < m$$

y que todos los menores de orden $r+1$ de la matriz jacobiana son idénticamente nulos en $U(x_0)$. Entonces, existe $\delta_0 > 0$ tal que las funciones $f_1(x), \dots, f_m(x)$ son funcionalmente dependientes en $B_{\delta_0}(x_0) \subseteq U(x_0)$.

Demostración. Sea $J_r(x)$ un menor de $Df(x)$ de orden r tal que $J_r(x_0) \neq 0$; por la continuidad de las derivadas $\frac{\partial f_i}{\partial x_j}$, $J_r(x) \neq 0$ en $x \in B_\rho(x_0)$ para un cierto $\rho > 0$, mientras que todos los menores de orden mayor que r son idénticamente nulos en $B_\rho(x_0)$. Recordando, si es necesario, las funciones f_1, \dots, f_m y las variables x_1, \dots, x_n podemos suponer sin pérdida de generalidad que $J_r(x)$ es el menor $r \times r$ que ocupa la esquina superior izquierda de $Df(x)$:

$$J_r(x) = \begin{vmatrix} \frac{\partial f_1}{\partial x_1} & \dots & \frac{\partial f_1}{\partial x_r} \\ \vdots & & \vdots \\ \frac{\partial f_r}{\partial x_1} & \dots & \frac{\partial f_r}{\partial x_r} \end{vmatrix}$$

Poniendo $x = (y, z)$, $y = (x_1, \dots, x_r)$, $z = (x_{r+1}, \dots, x_n)$, $f = (g, h)$, $g = (f_1, \dots, f_r)$, $h = (f_{r+1}, \dots, f_m)$, se tiene

$$Df(x) = \begin{pmatrix} D_y g(y, z) & D_z g(y, z) \\ D_y h(y, z) & D_z h(y, z) \end{pmatrix}$$

con la matriz $D_y g(y, z)$ no singular en $B_\rho(x_0)$ ya que $\det D_y g(y, z) = J_r(x) \neq 0$ si $x \in B_\rho(x_0)$.

Por el teorema de la función implícita, existen $\delta > 0$, $\varepsilon > 0$ tales que la ecuación en y, z, u ($y \in \mathbb{R}^r$, $z \in \mathbb{R}^{n-r}$, $u \in \mathbb{R}^r$)

$$g(y, z) = u$$

define, para (z, u) tales que $|z - z_0| < \delta$, $|u - u_0| < \delta$, donde $z_0 = (x_{r+1}^0, \dots, x_n^0)$, $u_0 = g(y_0, z_0) = (f_1(x_0), \dots, f_r(x_0))$, una única función de clase C^1

$$y = \phi(z, u)$$

que toma valores en $B_\varepsilon(y_0)$. Entonces, para (y, z) tales que $|y - y_0| < \varepsilon$ y $|z - z_0| < \delta$, se tiene

$$h(y, z) = h(\phi(z, u), z)$$

Si probamos que $H(z, u) = h(\phi(z, u), z)$ es independiente de z , o sea, de la forma $\varphi(u)$ en el conjunto $|z - z_0| < \delta$, $|u - u_0| < \delta$, entonces se tendrá

$$h(y, z) = \varphi(g(y, z))$$

para $|y - y_0| < \varepsilon$, $|z - z_0| < \delta$, es decir

$$\begin{cases} f_{r+1}(x) = \varphi_{r+1}(f_1(x), \dots, f_r(x)) \\ \vdots \\ f_m(x) = \varphi_m(f_1(x), \dots, f_r(x)) \end{cases}$$

para $x \in B_{\delta_0}(x_0)$ con $\delta_0 = \min\{\delta, \varepsilon\}$, lo que quiere decir que $f_{r+1}(x), \dots, f_m(x)$ son funcionalmente dependientes de $f_1(x), \dots, f_r(x)$ en $B_{\delta_0}(x_0)$.

Para probar que $H(z, u) = h(\phi(z, u), z)$ es independiente de z , calculamos su diferencial respecto a z por la regla de la cadena:

$$D_z H(z, u) = D_y h \cdot D_z \phi + D_z h$$

y aplicamos también dicha regla a la función compuesta idénticamente nula

$$\begin{aligned} (z, u) &\longmapsto g(\phi(z, u), z) - u = 0 : \\ D_y g \cdot D_z \phi + D_z g &= 0 \text{ (la matriz } 0 \text{ } r \times (m-r)) \end{aligned}$$

Pero, por la hipótesis del rango que hemos hecho, ha de existir una matriz P $(m-r) \times r$ tal que, en

$$Df(x) = \begin{pmatrix} D_y g(y, z) & D_z g(y, z) \\ D_y h(y, z) & D_z h(y, z) \end{pmatrix}$$

se tiene $D_y h(y, z) = P \cdot D_y g(y, z)$, $D_z h(y, z) = P \cdot D_z g(y, z)$, de lo que se deduce

$$D_z H(z, u) = D_y h \cdot D_z \phi + D_z h = P \cdot (D_y g \cdot D_z \phi + D_z g) = 0$$

y, de aquí, que $H(z, u)$ es independiente de z en el conjunto $|z - z_0| < \delta$, $|u - u_0| < \delta$ como queríamos demostrar. (Obsérvese: $r < m$, por hipótesis, y, naturalmente, $r \leq n$ porque la dimensión de $Df(x)$ es $m \times n$. Si $m \leq n$, o sea, si el número de funciones es menor o igual que el de variables, que es la situación usual, entonces $r < n$ y todos los argumentos anteriores se aplican tal cual se han expuesto. Si, por el contrario, $m > n$, entonces está garantizado el que $\text{rango} \left(\frac{\partial f_i}{\partial x_j}(x_0) \right) < m$ y, en el caso límite en que $r = n$, la demostración es trivial ya que la ecuación de partida es $g(x) = u$ (pues $x \equiv y$) y define (por el teorema de la función inversa, de hecho) $x = \phi(u)$, que, sustituida en $h(x)$, da $H(u) = h(\phi(u)) = \varphi(u)$ que, obviamente, sólo depende de u . Así, pues, en el caso de tener más funciones que variables, si el rango de $\left(\frac{\partial f_i}{\partial x_j}(x_0) \right)$ se mantiene constante en un cierto entorno de x_0 , hay con seguridad dependencia entre las funciones en un entorno $B_{\delta_0}(x_0)$ de x_0).

9.4 Teoremas de Lagrange y Kuhn-Tucker

Teorema 9.6 (Lagrange) Sean f, g_1, \dots, g_m , $m < n$, funciones de clase C^1 en un subconjunto abierto $S \subseteq \mathbb{R}^n$ con valores en \mathbb{R} . Supongamos que $x^* = (x_1^*, \dots, x_n^*)$ es un óptimo local de f en el conjunto

$$B = \{(x_1, \dots, x_n) \in S; g_j(x_1, \dots, x_n) = 0, j = 1, \dots, m\}$$

y supongamos también que los vectores $\nabla g_1(x^*), \dots, \nabla g_m(x^*)$ son linealmente independientes. Entonces existen constantes $\lambda_1^*, \dots, \lambda_m^*$ tales que

$$\nabla f(x^*) = \lambda_1^* \nabla g_1(x^*) + \dots + \lambda_m^* \nabla g_m(x^*) \quad (9.25)$$

Demostración: La demostración sigue los pasos de los de los argumentos dados en el apartado 6.2 utilizando ahora el teorema de la función implícita para funciones vectoriales a fin de resolver la ecuación $g(x) = 0$ siendo g la función vectorial

$$g(x) = \begin{pmatrix} g_1(x_1, \dots, x_n) \\ \vdots \\ g_m(x_1, \dots, x_n) \end{pmatrix}$$

Decir que $\nabla g_1(x^*), \dots, \nabla g_m(x^*)$ son linealmente independientes equivale a decir que el rango de la matriz $m \times n$

$$Dg(x^*) = \left(\frac{\partial g_j}{\partial x_i}(x^*) \right), i = 1, \dots, n, j = 1, \dots, m$$

es m (o sea, que es de rango máximo o completo). No hay pérdida de generalidad (reordenando las columnas de $Dg(x^*)$ si es necesario) si se supone que la submatriz formada por las m últimas columnas tiene determinante distinto de cero.

Denotemos por u al vector de \mathbb{R}^{n-m} formado por las $n-m$ primeras coordenadas de $x \in B$ y por v al vector de \mathbb{R}^m formado por las m últimas: $x = (u, v)$; en particular, $x^* = (u^*, v^*)$. La restricción es entonces

$$g(u, v) = 0$$

y se tiene, por lo anterior, que $D_u g(u^*, v^*) \in \mathcal{L}(\mathbb{R}^m)$ es invertible. Estamos, pues, en las hipótesis del teorema de la función implícita: existe una función $v = \varphi(u)$ tal que $g(u, \varphi(u)) = 0$ para $|u - u^*|$ suficientemente pequeño tal que $v^* = \varphi(u^*)$. Además $\varphi(u)$ es de clase C^1 en un entorno de u^* y se tiene

$$D\varphi(u^*) = - \overbrace{(D_v g(u^*, v^*))}^{m \times m}^{-1} \cdot \overbrace{D_u g(u^*, v^*)}^{m \times (n-m)} \in \mathcal{L}(\mathbb{R}^{n-m}, \mathbb{R}^m) \quad (9.26)$$

Puesto que la función

$$\mathbb{R}^n \ni u \mapsto f(u, \varphi(u)) \in \mathbb{R} \quad (9.27)$$

tiene un óptimo local en u^* habrá de ser, aplicando la regla de la cadena en (9.27),

$$\nabla_u f(u^*, v^*) + \nabla_v f(u^*, v^*) D\varphi(u^*) = 0 \in \mathbb{R}^{n-m} \quad (9.28)$$

de donde, utilizando (9.26),

$$\nabla_u f(u^*, v^*) - \nabla_v f(u^*, v^*) (D_v g(u^*, v^*))^{-1} D_u g(u^*, v^*) = 0 \quad (9.29)$$

Definamos ahora $\lambda^* = (\lambda_1^*, \dots, \lambda_m^*) \in \mathbb{R}^m$ por

$$\lambda^* = \overbrace{\nabla_v f(u^*, v^*)}^{1 \times m} \cdot \overbrace{(D_v g(u^*, v^*))^{-1}}^{m \times (n-m)} \quad (9.30)$$

Entonces

$$\nabla_v f(u^*, v^*) = \lambda^* D_v g(u^*, v^*) \quad (9.31)$$

por definición, y (9.29) toma la forma

$$\nabla_u f(u^*, v^*) = \lambda^* D_u g(u^*, v^*) \quad (9.32)$$

Explicitando (9.32) y (9.31) en términos de las variables $x = (x_1, \dots, x_n)$ originales se obtiene, respectivamente,

$$\left(\frac{\partial f}{\partial x_1}(x^*), \dots, \frac{\partial f}{\partial x_{n-m}}(x^*) \right) = (\lambda_1^*, \dots, \lambda_m^*) \begin{pmatrix} \frac{\partial g_1}{\partial x_1}(x^*) & \dots & \frac{\partial g_1}{\partial x_{n-m}}(x^*) \\ \vdots & \dots & \vdots \\ \frac{\partial g_m}{\partial x_1}(x^*) & \dots & \frac{\partial g_m}{\partial x_{n-m}}(x^*) \end{pmatrix} \quad (9.33)$$

$$\left(\frac{\partial f}{\partial x_{n-m+1}}(x^*), \dots, \frac{\partial f}{\partial x_n}(x^*) \right) = (\lambda_1^*, \dots, \lambda_m^*) \begin{pmatrix} \frac{\partial g_1}{\partial x_{n-m+1}}(x^*) & \dots & \frac{\partial g_1}{\partial x_n}(x^*) \\ \vdots & \dots & \vdots \\ \frac{\partial g_m}{\partial x_{n-m+1}}(x^*) & \dots & \frac{\partial g_m}{\partial x_n}(x^*) \end{pmatrix} \quad (9.34)$$

La reunión de (9.33) y (9.34) es la relación (9.25) que queríamos demostrar ($\nabla f(x^*) = \lambda^* Dg(x^*)$, en términos concisos).

A continuación, demostraremos a partir del teorema de Lagrange el resultado de Kuhn-Tucker que proporciona las condiciones necesarias de primer orden de los problemas de optimización con restricciones de desigualdad. El planteamiento general de un problema de este tipo era:

$$\begin{aligned} & \text{opt } f(x_1, \dots, x_n), \quad (x_1, \dots, x_n) \in S \subseteq \text{Dom } f \\ & \text{s.a.} \\ & \begin{cases} g_1(x_1, \dots, x_n) \leq 0 \\ g_2(x_1, \dots, x_n) \leq 0 \\ \vdots \\ g_m(x_1, \dots, x_n) \leq 0 \end{cases} \end{aligned} \quad (9.35)$$

Teorema 9.7 (Kuhn-Tucker) Sean f, g_1, \dots, g_m , funciones de clase C^1 en un subconjunto abierto $S \subseteq \mathbb{R}^n$ con valores en \mathbb{R} . Supongamos que $x^* = (x_1^*, \dots, x_n^*)$ es un óptimo local de f en el conjunto factible

$$B = \{x \in S; g_j(x) \leq 0, j = 1, \dots, m\} \quad (9.36)$$

o sea, una solución local del problema (9.35). Reordenando las funciones g_j si es necesario, podemos suponer que las restricciones de desigualdad que se saturan en x^* son $g_1(x^*) = 0, \dots, g_r(x^*) = 0$, con $r \leq m$. Pues bien, si los vectores $\{\nabla g_1(x^*), \dots, \nabla g_r(x^*)\}$ son linealmente independientes, entonces existen constantes $\lambda_1^*, \dots, \lambda_m^*$ tales que

$$\begin{cases} \nabla f(x^*) = \lambda_1^* \nabla g_1(x^*) + \dots + \lambda_m^* \nabla g_m(x^*) \\ y, \text{ para todo } j = 1, \dots, m, \quad \lambda_j^* g_j(x^*) = 0 \text{ y} \\ \lambda_j^* \leq 0 \text{ si } x^* \text{ es un mínimo o } \lambda_j^* \geq 0 \text{ si } x^* \text{ es un máximo.} \end{cases} \quad (9.37)$$

Demostración: Con la idea de aplicar el teorema de Lagrange, tratamos de convertir el problema (9.35) en un problema equivalente con restricciones de igualdad. Para ello, se introducen unas variables de holgura u_j^2 positivas que transforman cada restricción de desigualdad $g_j(x) \leq 0$ en una restricción de igualdad $g_j(x) + u_j^2 = 0$. Formulamos entonces el problema

$$\begin{aligned} & \text{opt } F(x, u), \quad (x, u) \in S \times \mathbb{R}^m \\ & \text{s.a. } G_j(x, u) = 0, \quad j = 1, \dots, m \end{aligned} \quad (9.38)$$

donde $u = (u_1, \dots, u_m) \in \mathbb{R}^m$, $F(x, u) \equiv f(x)$, $G_j(x, u) = g_j(x) + u_j^2$. Se trata de un problema de optimización en $n + m$ variables y m restricciones de igualdad. Si $x^* = (x_1^*, \dots, x_n^*)$ es una solución local del problema (9.35) en

la que se saturan las r primeras restricciones: $g_1(x^*) = 0, \dots, g_r(x^*) = 0$, entonces (x^*, u^*) con

$$u_1^* = \dots = u_r^* = 0; u_j^* = \sqrt{-g_j(x^*)}, j = r+1, \dots, m \quad (9.39)$$

es solución de (9.38) (y, recíprocamente, si (x^*, u^*) es solución de (9.38) y $u_1^* = \dots = u_r^* = 0, u_{r+1}^* \neq 0, \dots, u_m^* \neq 0$, entonces x^* es solución de (9.35) en la que sólo se saturan las r primeras restricciones). Además, si los vectores $\{\nabla g_1(x^*), \dots, \nabla g_r(x^*)\}$ son linealmente independientes, o sea, si la matriz jacobiana $Dg(x^*)$ —donde $g = (g_1, \dots, g_m)$ — tiene un menor formado con las r primeras filas que es no nulo, lo mismo ocurre con los vectores

$$\{\nabla G_1(x^*, u^*), \dots, \nabla G_m(x^*, u^*)\}$$

En efecto, construyendo la matriz jacobiana $DG(x^*, u^*)$ —donde $G = (G_1, \dots, G_m)$ — se observa que la matriz $m \times m$ construida con las $r \leq m$ columnas correspondientes al menor de $Dg(x^*)$ citado y las $m-r$ últimas, tiene determinante no nulo (por ser $u_j^* \neq 0, j = r+1, \dots, m$); por consiguiente, $DG(x^*, u^*)$ tiene rango máximo, o lo que es lo mismo, sus filas, los vectores $\{\nabla G_1(x^*, u^*), \dots, \nabla G_m(x^*, u^*)\}$, son linealmente independientes.

Estamos, por tanto, en las hipótesis del teorema de Lagrange para el problema (9.38) y podemos concluir que existen números $\lambda_1^*, \dots, \lambda_m^*$ tales que

$$\nabla F(x^*, u^*) = \lambda_1^* \nabla G_1(x^*, u^*) + \dots + \lambda_m^* \nabla G_m(x^*, u^*) \quad (9.40)$$

Podemos expresar esta condición en términos del lagrangiano

$$H(x, u, \lambda) = f(x) - \sum_{j=1}^m \lambda_j (g_j(x) + u_j^2) \quad (9.41)$$

del problema (9.38): (x^*, u^*, λ^*) ha de satisfacer

$$\frac{\partial H}{\partial x_i} = \frac{\partial f}{\partial x_i} - \sum_{j=1}^m \lambda_j \frac{\partial g_j}{\partial x_i} = 0, i = 1, \dots, n \quad (9.42)$$

$$\frac{\partial H}{\partial \lambda_j} = -(g_j(x) + u_j^2) = 0, j = 1, \dots, m \quad (9.43)$$

$$\frac{\partial H}{\partial u_j} = -2\lambda_j u_j = 0, j = 1, \dots, m \quad (9.44)$$

Las ecuaciones (9.43) son simplemente una reformulación de las restricciones $g_j(x) \leq 0$. Por otro lado, de (9.43) y (9.44) se deduce que

$$\lambda_j g_j(x) = 0, \quad j = 1, \dots, m \quad (9.45)$$

Hemos obtenido todas las conclusiones del teorema de Kuhn-Tucker salvo la relativa al signo de los multiplicadores. Para establecerla, consideramos el problema de minimización (los cambios a realizar para el de maximización son obvios). Veamos que $\lambda_1 \leq 0$; naturalmente, los argumentos para probar que $\lambda_j \leq 0, j = 2, \dots, r$ serán completamente análogos. Por la hipótesis de que los vectores $\{\nabla g_1(x^*), \dots, \nabla g_r(x^*)\}$ son linealmente independientes, podemos suponer sin pérdida de generalidad que

$$\frac{\partial(g_1, \dots, g_r)}{\partial(x_1, \dots, x_r)}(x^*) \neq 0 \quad (9.46)$$

Consideremos el siguiente sistema de r ecuaciones en las $r+1$ variables x_1, \dots, x_r, α :

$$\begin{aligned} g_1(x_1, \dots, x_r, x_{r+1}^*, \dots, x_n^*) + \alpha &= 0 \\ g_2(x_1, \dots, x_r, x_{r+1}^*, \dots, x_n^*) &= 0 \\ &\vdots \\ g_r(x_1, \dots, x_r, x_{r+1}^*, \dots, x_n^*) &= 0 \end{aligned} \quad (9.47)$$

$\{\alpha = 0, (x_1, \dots, x_r) = (x_1^*, \dots, x_r^*)\}$ es solución de este sistema. Entonces, por (9.46), el teorema de la función implícita garantiza que para α suficientemente pequeño, están definidas de modo único unas funciones

$$x_1 = \varphi_1(\alpha), \dots, x_r = \varphi_r(\alpha)$$

tales que $x_1^* = \varphi_1(0), \dots, x_r^* = \varphi_r(0)$ y

$$\begin{aligned} g_1(\varphi_1(\alpha), \dots, \varphi_r(\alpha), x_{r+1}^*, \dots, x_n^*) + \alpha &= 0 \\ g_j(\varphi_1(\alpha), \dots, \varphi_r(\alpha), x_{r+1}^*, \dots, x_n^*) &= 0, \quad j = 2, \dots, r \end{aligned} \quad (9.48)$$

Para valores $\alpha > 0$ suficientemente pequeños, se tiene

$$(\varphi_1(\alpha), \dots, \varphi_r(\alpha), x_{r+1}^*, \dots, x_n^*) \in B$$

ya que

$$\begin{aligned} g_1(\varphi_1(\alpha), \dots, \varphi_r(\alpha), x_{r+1}^*, \dots, x_n^*) &= -\alpha < 0 \\ g_j(\varphi_1(\alpha), \dots, \varphi_r(\alpha), x_{r+1}^*, \dots, x_n^*) &= 0, \quad j = 2, \dots, r \end{aligned}$$

y, para $j = r+1, \dots, m$, como $g_j(x^*) < 0$ y g_j y $\varphi_1, \dots, \varphi_r$ son funciones continuas se verificará

$$g_j(\varphi_1(\alpha), \dots, \varphi_r(\alpha), x_{r+1}^*, \dots, x_n^*) < 0$$

Entonces, por ser x^* un mínimo local de f relativo al conjunto B , se tendrá, para $\alpha > 0$ suficientemente pequeño,

$$\begin{aligned} 0 &\leq \frac{f(\varphi_1(\alpha), \dots, \varphi_r(\alpha), x_{r+1}^*, \dots, x_n^*) - f(x^*)}{\alpha} = \\ &= \frac{1}{\alpha} \sum_{i=1}^r \left[\frac{\partial f}{\partial x_i}(x^*) \cdot (\varphi_i(\alpha) - x_i^*) + (\varphi_i(\alpha) - x_i^*) \cdot \delta_i(\varphi(\alpha) - \bar{x}^*) \right] = \\ &\quad (\text{donde } \varphi(\alpha) - \bar{x}^* = (\varphi_1(\alpha) - x_1^*, \dots, \varphi_r(\alpha) - x_r^*)) \\ &= \sum_{i=1}^r \left[\frac{\partial f}{\partial x_i}(x^*) \cdot \frac{\varphi_i(\alpha) - \varphi_i(0)}{\alpha} + \frac{\varphi_i(\alpha) - \varphi_i(0)}{\alpha} \cdot \delta_i(\varphi(\alpha) - \bar{x}^*) \right] \\ &\xrightarrow{\alpha \rightarrow 0^+} \sum_{i=1}^r \left[\frac{\partial f}{\partial x_i}(x^*) \cdot \varphi_i'(0) + \varphi_i'(0) \cdot 0 \right] = \sum_{i=1}^r \frac{\partial f}{\partial x_i}(x^*) \cdot \varphi_i'(0) \end{aligned}$$

Pero, por (9.42) y (9.45),

$$\begin{aligned} \sum_{i=1}^r \frac{\partial f}{\partial x_i}(x^*) \cdot \varphi_i'(0) &= \sum_{i=1}^r \left[\sum_{j=1}^r \lambda_j \frac{\partial g_j}{\partial x_i}(x^*) \right] \cdot \varphi_i'(0) = \\ &= \sum_{j=1}^r \lambda_j \left[\sum_{i=1}^r \frac{\partial g_j}{\partial x_i}(x^*) \cdot \varphi_i'(0) \right] \end{aligned}$$

Derivando implícitamente en (9.48) se tiene

$$\begin{aligned} \sum_{i=1}^r \frac{\partial g_1}{\partial x_i}(x^*) \cdot \varphi_i'(0) &= -1 \\ \sum_{i=1}^r \frac{\partial g_j}{\partial x_i}(x^*) \cdot \varphi_i'(0) &= 0 \quad \text{para } j = 2, \dots, r \end{aligned}$$

con lo que

$$0 \leq \sum_{i=1}^r \frac{\partial f}{\partial x_i}(x^*) \cdot \varphi_i'(0) = -\lambda_1$$

es decir, $\lambda_1 \leq 0$, según queríamos probar.

9.5 Notas y complementos

1.- El teorema de la aplicación contractiva es válido en cualquier espacio métrico completo X (o sea, uno en el que toda sucesión de Cauchy es convergente). Basta sustituir en la demostración del apartado 9.1 la norma de \mathbb{R}^n por la distancia d en X , escribiendo $d(x, y)$ en lugar de $|x - y|$. Nótese, en particular, que la estructura de espacio vectorial de \mathbb{R}^n no ha intervenido para nada en la demostración. De la misma manera, el teorema de la contracción uniforme es válido siendo A un espacio métrico cualquiera y X un espacio métrico completo. Como se apuntaba al comienzo del capítulo, la utilización del teorema del punto fijo de Banach es uno de los caminos usuales para demostrar la existencia de soluciones de una ecuación diferencial (véase [6]).

2.- El teorema de la función implícita se puede extender a espacios de Banach con un enunciado prácticamente idéntico al del teorema 9.3:

Teorema 9.8 Sean X, Y, Z espacios de Banach, $\Omega \subseteq X \times Y$ un abierto, y sea $F: \Omega \rightarrow Z$ una función de clase C^k en Ω . Supongamos que el punto $(x_0, y_0) \in \Omega$ es tal que:

1. $F(x_0, y_0) = 0$.
2. $D_y F(x_0, y_0)$ es biyectiva y tiene inversa continua.

Entonces existen $\delta > 0, \varepsilon > 0$ y una función $f: B_\delta(x_0) \rightarrow B_\varepsilon(y_0)$ de clase C^k tales que

1. $f(x_0) = y_0$.
2. $F(x, f(x)) = 0$ para todo $x \in B_\delta(x_0)$.
3. Para cada $x \in B_\delta(x_0)$, $y = f(x)$ es la única solución de la ecuación $F(x, y) = 0$ que pertenece a $B_\varepsilon(y_0)$.
4. Además, las diferenciales sucesivas de f en $x \in B_\delta(x_0)$ se pueden obtener por diferenciación implícita. En particular,

$$Df(x) = -(D_y F(x, f(x)))^{-1} D_x F(x, f(x))$$

La noción de diferencial parcial es la misma que la que hemos dado en el apartado 9.2 (valen las proposiciones 9.1 y 9.2) y la demostración del teorema es, casi palabra por palabra, la que hemos visto para los espacios \mathbb{R}^n (únicamente hay que sustituir los argumentos en los que hemos utilizado el determinante de una matriz). El teorema de la función implícita en espacios de Banach es muy útil en el estudio avanzado de ecuaciones diferenciales y otras ecuaciones funcionales (véase [6]).

Parte III

El Cálculo Diferencial en la Economía

Capítulo 10

Nociones de economía marginalista

En esta parte final del libro aplicaremos los conceptos y métodos de los capítulos anteriores a ciertos aspectos de la teoría económica: veremos cómo el cálculo diferencial permite desarrollar una teoría muy elegante sobre el comportamiento de consumidores y productores en una economía de mercado, partiendo de unas pocas hipótesis que, aunque ciertamente simplificadoras, no dejan de ser bastante razonables.

En el presente capítulo introduciremos los conceptos fundamentales acerca de la noción de *valor de un bien* y los condicionamientos a que se ven sometidos tanto los productores –generadores de la *oferta*– como los consumidores del mismo –generadores de la *demanda*–, todo ello desde el punto de vista llamado *marginalista*. En el apartado 10.2 se comentan ciertas peculiaridades que acompañan a la aplicación del cálculo diferencial a la economía, como son las nociones de *magnitud marginal* y de *elasticidad*. El último apartado contiene unos ejemplos numéricos sencillos en los que las funciones de oferta y demanda introducidas en el primer apartado se pueden calcular explícitamente, dejando para el capítulo siguiente el análisis general de las mismas.

10.1 La teoría marginalista

Es conocida la gran dificultad que entraña la formalización matemática de la economía (“la ciencia desastre”, como la llaman muchos de sus cultivadores). Una de las grandes diferencias entre las ciencias sociales y las físicas es que en aquéllas los objetos de estudio son *agentes activos* (valga la redundancia) capaces de *tomar decisiones* que afectan al comportamiento del propio sistema; en el caso de la ciencia económica, y según la definición ya consagrada

de Lionel Robbins, los agentes económicos eligen en un universo de *bienes y servicios escasos susceptibles de usos alternativos*. Otra diferencia importante entre ambos tipos de ciencias está en la dificultad de establecer *jerarquías u órdenes de magnitud* en el fenómeno social: los escalones tan nítidos que se presentan en las ciencias físicas y biológicas (los niveles subatómico, atómico, molecular, celular,...) y que hacen que el análisis del fenómeno se pueda repartir práctica y metodológicamente entre ellos, desaparecen prácticamente en la ciencia económica con escasas excepciones, como la tradicional, difícil y discutida distinción entre macroeconomía y microeconomía. En la práctica, la escala de actuación del agente económico típico no es, en general, ni tan minúscula y local como la de un átomo ni tan global como un organismo entero.

La teoría del valor

Los economistas clásicos trataron de explicar el *precio* de un producto como resultado de su *coste* de producción, de forma que, por ejemplo, si producir A cuesta el doble que producir B pero su precio de venta no es el doble del de B, los productores tenderán a dejar de fabricar A para dedicarse a fabricar B, a consecuencia de lo cual se producirán menos unidades de A y más de B. Se debe, pues, concluir que la única forma de que se observen *precios y cantidades estables* en la economía es que se dé una proporcionalidad perfecta entre costes de producción y precios de venta.

A pesar de las muchas y muy interesantes conclusiones que pueden sacarse de este modelo tan simplificado, los fundadores de la economía moderna descubrieron pronto que el coste de producción de una mercancía es *sólo una parte* del problema de la determinación de su precio de mercado. El hecho de que los bienes tengan un valor y un precio se debe, en última instancia, a su *carácter limitado*, tanto en términos absolutos (si en su producción intervienen recursos no renovables) como en términos relativos (pues los recursos en él empleados podrían haberse utilizado en la fabricación de otros productos). Vemos, pues, que el coste monetario de un producto puede no medir su verdadero coste social, ya que cada proceso de producción constituye una cadena de decisiones de empleo de recursos que bien podrían haberse utilizado en otra cosa. El coste de producción, pues, debería reflejar también el coste de renunciar a la oportunidad de emplear los mismos recursos para otros fines.

Pero no es ésta la única dificultad. Como ya observaron los economistas clásicos, el precio de un objeto único en su especie, como un incunable, no guarda relación a su coste de producción, sino a su *demanda*, la cual está, a su vez, en función de la *utilidad* que su posesión reporta a los coleccionistas del mismo. Y, sin embargo, la consideración de la utilidad en sí misma como factor determinante del precio es también problemática: siendo el pan un

producto vital, su utilidad debería ser mucho mayor que la de un Rembrandt o un diamante y, sin embargo, su precio de mercado es muy inferior.

La teoría de la *utilidad marginal* del último tercio del siglo XIX aborda y resuelve esta paradoja. La *teoría marginalista* parte de la idea de que el precio de un producto puede no reflejar directamente su "valor" medido en términos de la cantidad de recursos en él invertidos (coste de producción) o su "valor intrínseco" entendido como la utilidad que reporta a sus compradores; pero las variaciones en estas magnitudes absolutas (producción y utilidad) sí se relacionan directamente con las variaciones en el precio: no interesa tanto hablar de la "utilidad del pan" en términos absolutos, sino de la *diferencia de utilidad* que reporta a una persona el disponer de 250 g. de pan en vez de 240, diferencia que se conoce como la *utilidad marginal* que reportan esos últimos 10 g. de pan. Según los teóricos marginalistas, si esta utilidad marginal supera al precio de esos 10 g. de pan, el consumidor estará interesado en su adquisición, mientras que, en caso contrario, preferirá utilizar ese dinero en otras adquisiciones que le reporten mayor utilidad marginal, como, por ejemplo, mantequilla para untar el pan. Como es lógico, este análisis exige suponer que el producto final se puede expandir y contraer arbitrariamente, por lo que no es aplicable a "incunables" o productos únicos.

El consumidor, en definitiva, procura la consecución de unos *objetivos de consumo* (necesidades, gustos...) que le reportan unas ciertas utilidades, viéndose sometido a ciertas restricciones externas: los precios de los bienes y servicios que desea adquirir, su propia renta disponible, las restricciones institucionales (derechos de propiedad, reglas de intercambio) y las restricciones tecnológicas. El problema es que las preferencias del consumidor son mucho más difíciles de cuantificar que los precios, rentas, tasas, cuotas... a los que se enfrenta. La clave del método marginalista consiste en estudiar cómo se ven afectados los *cambios* en el comportamiento individual por los cambios en sus oportunidades de elección, es decir, en las *restricciones* a que está sometido: aun siendo quizá imposible asignar variables medibles a las preferencias, puede ser posible predecir los cambios que éstas experimentan como respuesta a los cambios que se producen en las restricciones, sobre todo, en precios y rentas. Digamos también que puede llevarse a cabo un análisis marginalista muy similar para los *productores*, es decir, los agentes económicos que producen y llevan los productos al mercado, y que pretenden maximizar su beneficio (= ventas - costes) hallándose también sometidos a *restricciones* de tipo presupuestario, tecnológico, etc.

Las funciones de oferta y demanda

Una vez asumido este paradigma marginalista, en el mercado de un bien o servicio los productores y consumidores del mismo reaccionan ante las varia-

ciones de su precio siguiendo unas pautas reconocibles (funciones de oferta y demanda, respectivamente). Estos comportamientos correctores hacen que si el precio vigente se encuentra en una situación de desequilibrio (por ejemplo, demasiado bajo para los productores o demasiado alto para los consumidores), éste tienda a modificarse en el sentido de compensar ese desequilibrio.

NOTA. En la teoría económica clásica este proceso se considera "automático", en el sentido de que no lo lleva a cabo ningún "regulador", sino, a lo sumo, un "anotador de ofertas" o "secretario del mercado", como en una subasta. No hay, sin embargo, acuerdo general sobre la descripción precisa del mecanismo de ajuste del precio. Encontramos aquí una situación muy frecuente en la teoría económica: hubo un consenso bastante general acerca de los modelos matemáticos que dan los valores teóricos de equilibrio (modelos estáticos) junto con una dificultad extrema para describir el proceso por el que se llega al equilibrio (modelos dinámicos).

La forma explícita en que productores y consumidores responden a las variaciones de precios depende de sus posibilidades y tecnologías de producción en el primer caso (sintetizados en la llamada *función de producción*), de sus necesidades y preferencias en el segundo (reunidos en la llamada *función de utilidad*) y de las *restricciones* a las que unos y otros se ven sometidos. En el capítulo siguiente se verá cómo se generan estas funciones a partir de hipótesis muy sencillas sobre la *racionalidad* del comportamiento de los agentes económicos. Sin embargo, una parte importante de la teoría de la determinación del precio de equilibrio se puede abordar tomando dichas funciones como *datos a priori*: por ejemplo, mediante mediciones estadísticas. Así, la *función de oferta* $S(p)$ expresa la cantidad de producto, medido en unidades físicas, que los productores están dispuestos a lanzar al mercado supuesto que se venda al precio unitario p . En condiciones normales, y por razones fáciles de entender, se trata de una función positiva, *creciente* (a mayor precio mayor beneficio y por tanto mayor producción) y con un umbral mínimo dado por los costes fijos de producción (instalaciones, máquinas, mantenimiento...). Por otra parte, la *función de demanda* $D(p)$ representa la cantidad del bien o servicio (también medido en unidades físicas) que el conjunto de los consumidores estaría dispuesto a comprar al precio unitario p . En los bienes para los que existen sustitutos (mantequilla y margarina, por ejemplo), un aumento del precio unitario provoca un desplazamiento del consumo hacia los bienes o servicios sustitutivos, lo que motiva un descenso en la demanda del bien en cuestión: $D(p)$ es, pues, una función *decreciente* de p . Si además el bien no produce saciedad ("cuanto más, mejor") y se puede almacenar fácilmente sin deteriorarse (como el dinero) parece lógico suponer que $D(p) \rightarrow \infty$ cuando $p \rightarrow 0$.

Si para un valor de p se tiene que $S(p) > D(p)$, quedarán en el mercado $S(p) - D(p)$ artículos sin vender: ante los gastos que esto genera a los productores, éstos preferirán marcar precios inferiores con tal de atraerse nuevos

compradores y vender toda la producción. Recíprocamente, si $S(p) < D(p)$, la existencia de demanda sin cubrir permitirá a los productores elevar sus precios sin perder por ello clientela. En definitiva, el precio p^* para el cual se equilibran oferta y demanda, $S(p^*) = D(p^*)$, tendrá la propiedad de que *todo* lo producido encontrará comprador y, si las condiciones se mantienen igual, el precio no tenderá ni a subir ni a bajar: p^* será, pues, el *precio de equilibrio* del producto en cuestión (figura 10.1).



Figura 10.1 Precio de equilibrio

Vemos así la importancia de entender bien el comportamiento de las funciones de oferta y demanda y, en particular, sus variaciones respecto del precio. Para ello, y salvo en casos muy simples, el cálculo diferencial es la herramienta más poderosa, si bien con ciertas adaptaciones al ámbito económico, cuestión que discutimos a continuación.

10.2 Magnitudes marginales y elasticidades

Por razones históricas, las derivadas de las magnitudes económicas se suelen denominar con el calificativo de "magnitud marginal asociada", y se interpretan, en general, en términos *finitos*, es decir, no diferenciales: la magnitud marginal MZ asociada a una magnitud económica Z dependiente de una variable x se define formalmente como $Z'(x)$ pero se interpreta como el incremento generado en Z al incrementarse x en una unidad. En símbolos, $MZ(x) = Z'(x) \simeq Z(x+1) - Z(x)$, que no es otra cosa que el *coeficiente incremental* $\Delta Z/\Delta x$ cuando $\Delta x = 1$. Dado el tamaño de los números que habitualmente se manejan en economía, se puede considerar que un incremento unitario es realmente "infinitesimal" en relación con dichas cantidades, o, dicho de forma más precisa, que la aproximación de primer orden $S(x+\Delta x) - S(x) \simeq S'(x)\Delta x$ es suficientemente buena cuando $\Delta x = 1$.

NOTA. La mayoría de las nociones marginales se entienden mucho mejor cuando se conciben en estos términos "finitos" que cuando se interpretan exactamente como derivadas, y se puede llegar muy lejos en la teoría prescindiendo totalmente de estas últimas. Sin

embargo, el cálculo diferencial es una herramienta de tal potencia que los teóricos no tardaron en emplearla a fondo con todas sus consecuencias.

Veamos algunos ejemplos:

- Si $U(q)$ representa la utilidad (medida en términos monetarios) que proporciona al consumidor la posesión de q unidades de un bien, la *utilidad marginal* asociada será $U'(q)$, que representa (aproximadamente) el incremento de utilidad que proporciona la posesión de una unidad más ($q + 1$) del bien en cuestión. Según los teóricos marginalistas, si p es el precio de mercado de un bien y lo consideramos como un dato, un consumidor *demandará* el número q^* de unidades del mismo que haga que su utilidad marginal $MU(q^*) = U'(q^*)$ coincida con p . Esta relación entre p y q^* será, pues, la *función de demanda* de ese consumidor.
- Si $C(q)$ representa el coste de producir q unidades de un bien, el *coste marginal* asociado $MC(q) = C'(q)$ es (aproximadamente) el coste de producir una unidad adicional del bien, es decir, $C(q+1) - C(q)$. Según la teoría marginalista, una empresa con financiación ilimitada fabricará q^* unidades del producto al precio de mercado (dado *a priori*) p siempre que q^* maximice el *beneficio* $B(q) = pq - C(q)$, lo cual tendrá lugar cuando $B'(q^*) = 0$, es decir, cuando $p - C'(q^*) = 0$, o bien, $p = C'(q^*)$: la fábrica seguirá aumentando su producción hasta que la última unidad fabricada cueste exactamente p . Véase la relación entre coste de producción y precio de mercado que comentábamos en el primer apartado, pero se trata ahora del *coste marginal*.
- La posibilidad de resolver la ecuación $p = C'(q)$ en función de q dependerá de las propiedades de la función $C''(q)$. Si ésta no se anula, la teoría de funciones implícitas garantiza que, en efecto, para cada p de un cierto intervalo existe un *único* q tal que $p = C'(q)$. Si llamamos S a la relación funcional así establecida, $q = S(p)$, ésta representa precisamente la *función de oferta* de la empresa fabricante del producto. Su derivada $S'(p)$ se llama tradicionalmente *oferta marginal*, y representa el incremento de unidades del bien que la fábrica produciría si el precio unitario aumentase en una unidad. Análogamente tendríamos la noción de *demanda marginal*.

Elasticidad

Otra peculiaridad del cálculo diferencial en sus aplicaciones económicas es el frecuente uso de la noción de *elasticidad*, variante de la idea de derivada. Si en

la definición de ésta se tienen en cuenta los cocientes incrementales de las variables en relación funcional, en el caso de la elasticidad lo que se considera son los cocientes de sus *incrementos relativos* o porcentuales: es mucho más significativo decir, por ejemplo, que "la demanda de vivienda bajó un 60% cuando los precios del suelo se duplicaron" que afirmar que "se vendieron 180.000 viviendas menos que en el período precedente porque el metro cuadrado de suelo edificable subió 100.000 pts".

Consideremos, en general, dos variables x , y relacionadas funcionalmente por $y = f(x)$. Si, en lugar de estudiar los cocientes de incrementos absolutos $\Delta y / \Delta x = (f(x + \Delta x) - f(x)) / \Delta x$ nos fijamos en los cocientes de los *incrementos relativos*

$$\frac{\Delta y / y}{\Delta x / x}$$

y pasamos al límite cuando $\Delta x \rightarrow 0$, obtenemos el concepto de *elasticidad de y respecto de x* . En términos finitos, mucho más fáciles de entender e interpretar, si suponemos que Δx es la centésima parte de x , con lo que el incremento relativo $\Delta x / x$ es el 1%, la elasticidad de y representaría el incremento relativo (o sea, en términos porcentuales) de y en respuesta al incremento del 1% experimentado por x . Si hacemos ahora tender Δx a 0, el límite del cociente de incrementos relativos es, por definición, la *elasticidad de y* (o de f). Para distinguir este concepto del anterior, de la misma forma que se distingue entre *velocidad media* y *velocidad instantánea*, se llama a veces "elasticidad-arco" al cociente incremental correspondiente a un valor finito (y dado) de Δx , y "elasticidad-punto" o "puntual" al límite antes mencionado.

Por tanto, si llamamos E ó $E(f)$ a la elasticidad (puntual), tendremos:

$$\begin{aligned} E(f) &= E_x(f) \stackrel{\text{def}}{=} \lim_{h \rightarrow 0} \frac{\frac{f(x+h) - f(x)}{h}}{\frac{h}{x}} = \lim_{h \rightarrow 0} \frac{x}{f(x)} \frac{f(x+h) - f(x)}{h} = \\ &= \frac{x f'(x)}{f(x)} = \frac{xy'}{y} \end{aligned}$$

donde, naturalmente, se sobreentiende que $f(x) \neq 0$.

Por ejemplo, si la demanda de un bien es $q = D(p) = 100 - 25p$, la elasticidad de la demanda (respecto del precio) sería

$$E_p(D) = \frac{p D'(p)}{D(p)} = \frac{-25p}{100 - 25p} = 1 - \frac{100}{100 - 25p}$$

NOTA. Como $D(p)$ es decreciente, $E(D)$ es negativa, razón por la cual a veces se cambia el signo para que esta cantidad sea positiva, definiendo la elasticidad como $-E$. Nosotros no seguiremos aquí este convenio.

Se dice que la demanda es *inelástica* (en p^*) si $|E(D)(p^*)| < 1$, y *elástica* cuando $|E(D)(p^*)| > 1$. En el ejemplo anterior, $E = 0$ para $p = 0$, $|E| = 1$ ($E = -1$) para $p = 2$ (por lo que la demanda es inelástica para $0 < p < 2$ y elástica para $p > 2$) y $E \rightarrow -\infty$ cuando $p \rightarrow 4$. En efecto, si pasamos de $p = 3,9$ a $p = 3,8$, el incremento absoluto de p es $\Delta p = -0,1$, su incremento relativo es $\Delta p/p = -0,1/3,9 \simeq -0,25\%$, y los valores correspondientes de la demanda son $D(3,9) = 2,5$, $D(3,8) = 5$, $\Delta D = 2,5$, $\Delta D/D = 1 = 100\%$. Es decir, un descenso del 0,25% en el precio del producto se ha traducido en un incremento del 100% en la demanda del mismo. Obsérvese que, en términos absolutos, el incremento de demanda es muy pequeño: $\Delta D = 2,5$ unidades del bien; sin embargo, el paso de 2,5 a 5 unidades no deja de ser un incremento del 100%.

Elasticidad y transformación logarítmica

Observemos que

$$E(f) = \frac{xf'(x)}{f(x)} = \frac{f'(x)/f(x)}{1/x} = \frac{[\log f(x)]'}{[\log x]'}$$

supuesto que tanto x como $f(x)$ sean positivos. Esto indica que, si llamamos $X = \log x$ y expresamos $f(x)$ en términos de X (basta escribir $f(x) = f(e^X)$), entonces $E(f)$ es la *derivada ordinaria* de $\log f(e^X)$ como función de X , ya que

$$\frac{d}{dX} \log(f(e^X)) = \frac{e^X f'(e^X)}{f(e^X)} = \frac{xf'(x)}{f(x)} = E(f)$$

El formalismo de estas operaciones no debe ocultarnos el fondo de la cuestión: si incrementamos en h unidades el logaritmo del número N , obtenemos

$$\log N + h = \log N + \log(e^h) = \log(Ne^h)$$

lo que nos dice que un incremento *absoluto* de h unidades en $\log N$ se corresponde con el logaritmo de un incremento *relativo* del $(e^h - 1)\%$ en N . Dicho de otra manera, si en lugar de considerar unas magnitudes dadas nos fijamos en sus logaritmos, *los incrementos absolutos en los logaritmos corresponden a incrementos relativos en las magnitudes de partida*.

Todavía más: Si en un sistema de ejes cartesianos representamos, no las variables x e y , sino sus *logaritmos* $X = \log x$, $Y = \log y$, la curva $y = f(x)$ se transforma en la curva $Y = \log y = \log f(x) = \log f(e^X)$, que llamaremos $g(X)$. Pues bien, la *pendiente* de esta curva transformada $Y = g(X)$ (su *derivada*) es precisamente la *elasticidad* de y respecto de x .

Por ejemplo, la función de demanda anterior escrita en las variables $Q = \log q$, $P = \log p$ (y entonces $p = e^P$) sería:

$$Q = \log q = \log D(p) = \log(100 - 25p) = \log(100 - 25e^P)$$

y $E(D)$ se podría calcular como dQ/dP , y visualizar como la pendiente de la curva $Q = \log(100 - 25e^P)$ en el plano (P, Q) (véase la figura 10.2).

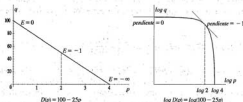


Figura 10.2

Esta transformación *logarítmica* se visualiza mucho más fácilmente usando *papel logarítmico*, es decir, papel milimetrado en el que, en los ejes, en vez de los valores reales de las magnitudes, se miden sus logaritmos.

Así como las funciones lineales son las que tienen derivada constante, las *funciones de elasticidad constante* deberán tener la misma propiedad tras efectuar la transformación logarítmica, es decir: $\log y = a + b \log x$, o sea, $y = Ax^b$, donde $A = e^a > 0$. Serán crecientes si $b > 0$ y decrecientes si $b < 0$. Las curvas de demanda de elasticidad constante se llaman *curvas marshallianas*, en honor al economista inglés Alfred Marshall, introductor de la noción de elasticidad en economía.

En funciones de varias variables se definen las *elasticidades parciales* de la manera obvia, y, por ejemplo, para que $z = f(x, y)$ tenga elasticidades parciales constantes hace falta que $\log z$, expresada en términos de $\log x, \log y$, tenga *derivadas parciales* constantes, es decir que $\log z = a \log x + b \log y + c$. O sea: $z = Kx^ay^b$, donde $K = e^c$. Encontramos así las famosas funciones de producción de Cobb-Douglas, de las que se habló en el Capítulo 1, y que se caracterizan por la propiedad de que un incremento relativo del $r\%$ en uno de los factores, manteniendo constantes los demás, genera, *bajo cualquier situación inicial* (x, y) , el mismo incremento relativo en la producción.

Usos de la elasticidad

Las elasticidades se usan tan frecuentemente en la teoría económica por una serie de razones: en un sentido general, el uso, tanto práctico como concep-

tual, de incrementos relativos es muy natural en economía porque, entre otros motivos, estos son independientes de la unidad de medida (física, monetaria...) utilizada.

Más específicamente, la elasticidad aparece también de manera natural al analizar ingresos o costes totales en función de sus precios unitarios. Por ejemplo, si los consumidores demandan (y adquieren) $D(p)$ unidades de un producto al precio unitario p , el importe total $V(p)$ de las ventas de ese producto al precio p será $V(p) = pD(p)$, cuya variación respecto de p vendrá dada por

$$V'(p) = pD'(p) + D(p) = D(p) \left(p \frac{D'(p)}{D(p)} + 1 \right) = D(p) (E(p) + 1)$$

donde $E(p) = E(D)(p) = pD'(p)/D(p)$ es, como sabemos, la elasticidad de D en p . Por tanto, siendo $D(p) > 0$, vemos que $V(p)$ será creciente en el intervalo de precios en que $E + 1 > 0$, que es el correspondiente a la demanda inelástica $0 \leq |E| < 1$, y será decreciente cuando $E + 1 < 0$, es decir, cuando $|E| > 1$ (demanda elástica). Si existe un precio p^* en el que la elasticidad es -1 y la demanda pasa en él de inelástica a elástica, dicho valor es el que maximizará el volumen total de ventas del producto en cuestión.

Por ejemplo, la demanda de bienes agrícolas suele ser inelástica, y un "cosechón" disminuye los precios de tal forma que el montante total de ventas es menor que en un año de sequía (naturalmente, si no se tienen en cuenta otros factores, como posibilidad de almacenamiento, etc.). O bien, la OPEP puede tener interés en reducir la producción de crudo, aumentando así tanto el precio como el volumen de ventas: la demanda de petróleo es ciertamente inelástica. En cambio, una maniobra de reducción de producción en un producto industrial de demanda elástica (quizá por tener muchos sustitutos) puede tener el efecto contrario. Es el valor de la elasticidad, no el de la pendiente, el que decide en cada uno de estos casos. En la figura 10.3, el aumento del precio produce un montante de ventas pq menor (izquierda), igual (centro) o mayor (derecha) que el de partida.

Para un completo tratamiento del tema de la elasticidad y, en general, de las aplicaciones del cálculo diferencial a la economía, pueden consultarse [9] y [11].

10.3 Oferta y demanda: ejemplos

Las funciones de oferta y demanda son muy intuitivas y constituyen la base del análisis económico en una primera aproximación, pero no es evidente cómo se pueden calcular, o al menos estimar, en términos de los elementos observables de una economía dada, de la tecnología que emplean sus empresas productoras

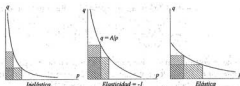


Figura 10.3

de bienes y servicios o de las preferencias y restricciones de los consumidores que los demandan. En el siguiente capítulo veremos cómo se aborda este problema bajo hipótesis bastante generales; aquí consideraremos simplemente unos ejemplos introductorios.

Costes y función de producción

La función de costes totales $C(q)$ antes mencionada se obtendría eligiendo, de entre todas las posibles estrategias de producción que llevan a la fabricación de q unidades de producto, aquella que requiera el *mínimo* coste.

EJEMPLO 1. Consideremos una empresa cuyo proceso productivo emplea x piezas de maquinaria operadas por y trabajadores, obteniendo como resultado $f(x, y) = 3 \log x + \log y$ unidades de un cierto producto, y supongamos que opera en régimen de competencia perfecta, es decir, que tanto el precio unitario p del producto final como los precios unitarios π_x , π_y de los factores de producción están dados, y la empresa no puede hacer nada por modificarlos. ¿Cuál sería el modo más racional de producir una cantidad dada q de producto, con independencia de que realmente interese producir esas q unidades, es decir, de que se puedan vender?

La empresa elegiría los valores $x^* \geq 0$, $y^* \geq 0$ que *minimizan* el *coste total* $\pi_x x + \pi_y y$ bajo la restricción de producir *al menos* q unidades, es decir, $f(x, y) \geq q$. Usualmente esta restricción se satisface con *igualdad*, es decir, $f(x, y) = q$, que es lo que supondremos ahora, dejando que el lector reflexione sobre el caso general.

Se trata, pues de *minimizar* $\pi_x x + \pi_y y$ *condicionado a* $3 \log x + \log y = q$. Por brevedad, procedamos por el método directo y despejemos y de esta última igualdad, obteniendo $y = e^{q-3 \log x}$. Con esta combinación de factores es seguro que se producirán q unidades de producto, y se trata ahora de elegir x para que el coste sea mínimo. El coste, que llamaremos $K(x)$ cuando lo expresemos como función del factor x , será $K(x) = \pi_x x + \pi_y e^{q-3 \log x}$, función estrictamente

convexa cuya derivada, $K'(x) = \pi_x - 3\pi_y e^q x^{-4}$ se anula exclusivamente para

$$x^* = \left(\frac{3\pi_y}{\pi_x} e^q \right)^{1/4} = \left(\frac{3\pi_y}{\pi_x} \right)^{1/4} e^{q/4}$$

con lo que el coste mínimo asociado a la producción de q unidades, que designaremos $C(q)$ cuando lo expresamos como función de q , es el que corresponde a la combinación de factores x^* e $y^* = e^q (x^*)^{-3}$, y vale:

$$\begin{aligned} C(q) &= \pi_x x^* + \pi_y e^q (x^*)^{-3} = \pi_x \left(\frac{3\pi_y}{\pi_x} e^q \right)^{1/4} + \pi_y e^q \left(\frac{3\pi_y}{\pi_x} e^q \right)^{-3/4} = \\ &= M \pi_x^{3/4} \pi_y^{1/4} e^{q/4} \end{aligned}$$

donde $M = 3^{1/4} + 3^{-3/4} \simeq 1.7548$.

El coste marginal $C'(q)$ es igual a $(M \pi_x^{3/4} \pi_y^{1/4} / 4) e^{q/4}$. Para maximizar su beneficio $B(q) = pq - C(q)$, la empresa producirá el número q^* de unidades que haga $B'(q^*) = p - C'(q^*) = 0$, es decir, que iguale coste marginal y precio de venta. En nuestro caso,

$$q = 4 \log \frac{4p}{M \pi_x^{3/4} \pi_y^{1/4}} \stackrel{\text{def}}{=} S(p)$$

que será, pues, la *función de oferta* de la empresa, que establece la cantidad q de producto que la empresa fabricaría si estuviese segura de que el precio de mercado de éste fuese p y de que, a ese precio, podría vender toda su producción q . Además, las expresiones de x^* e y^* nos indican las cantidades de los factores de producción que la empresa estaría dispuesta a adquirir, en función de sus precios π_x , π_y y de la cantidad a producir, q , que, a su vez, está en función de p a través de la función de oferta. Operando, se obtiene:

$$\begin{aligned} x^* &= \left(\frac{3\pi_y}{\pi_x} \right)^{1/4} e^{q/4} = \left(\frac{3\pi_y}{\pi_x} \right)^{1/4} \frac{4p}{M \pi_x^{3/4} \pi_y^{1/4}} = \frac{3p}{\pi_x}, \\ y^* &= e^q (x^*)^{-3} = \left(\frac{4p}{M \pi_x^{3/4} \pi_y^{1/4}} \right)^4 \left(\frac{3p}{\pi_x} \right)^{-3} = \frac{p}{\pi_y} \end{aligned}$$

que representan las *funciones de demanda de los factores*, expresadas en función de los precios de dichos factores y del precio del producto final. Vemos que se trata de funciones de demanda de elasticidad constante -1 , por lo que los cambios en los precios de los factores no influirán en la factura final que los suministradores le pasarán a la empresa. En este caso particular, además, x^* no depende de π_y , ni depende y^* de π_x , pero esto no es así en general.

Demanda y función de utilidad

Este método directo de análisis es también aplicable al caso del *consumidor* que, deseando consumir dos bienes x , y a precios unitarios p_x , p_y , elige la combinación de ambos que considere más deseable *dentro de sus posibilidades*, es decir, sin gastar más que el presupuesto r del que disponga para este objetivo. Así planteada, esta cuestión parece mucho más "subjetiva" que la anterior, dado que la función de producción es un *dato objetivo* para cada empresa; aun así, los economistas del siglo pasado trataron de establecer un paralelo completo entre ambas situaciones, suponiendo la existencia de una "función de utilidad" $u = f(x, y)$ que midiera el *grado de satisfacción* que produce el disfrute de x unidades del primer bien, junto con y unidades del segundo, y que jugaría el papel de la función de producción si se tratase de una empresa.

EJEMPLO 2. Si $f(x, y) = 3 \log x + \log y$ fuese la utilidad que le reporta a un consumidor la adquisición de x e y unidades de cada bien, éste se plantearía el problema de maximizar su utilidad $f(x, y)$, *condicionada a la restricción presupuestaria* $p_x x + p_y y = r$.

NOTA. Véase la diferencia entre este problema de optimización y el anterior, de minimizar $p_x x + p_y y$ condicionado a $f(x, y) = q$. Este es un ejemplo de dualidad entre problemas de optimización, cuestión muy importante en la que no podemos entrar ahora. El caso más antiguo de esta dualidad se da con el llamado "problema isoperimétrico": hallar la figura plana que, con un perímetro dado P encierre mayor área. Su dual sería: hallar la figura plana que, con un área dada A , tenga menor perímetro. La solución, en ambos casos, es el círculo.

Procediendo, como antes, por el método directo, despejamos $y = r/p_y - (p_x/p_y)x$ y sustituimos en la función de utilidad, obteniendo $F(x) \stackrel{\text{def}}{=} 3 \log x + \log (r/p_y - (p_x/p_y)x)$, función estrictamente cóncava cuya derivada

$$F'(x) = \frac{3}{x} + \frac{p_x/p_y}{(p_x/p_y)x - r/p_y}$$

se anula exclusivamente para $x^* = 3r/(4p_x)$, valor que, junto con su asociado $y^* = r/p_y - (p_x/p_y)(3r/(4p_x)) = r/(4p_y)$ representan la *elección óptima del consumidor* dados los precios p_x , p_y y el presupuesto r (compárense con las expresiones obtenidas arriba para la demanda de factores de producción). Si suponemos fijos r y p_y , por ejemplo, la expresión $x^* = 3r/(4p_x)$ no es otra cosa que la *función de demanda* del producto x en función del precio p_x del producto: la función de demanda $D(p_x)$ del análisis económico elemental.

Si, en vez de usar $f(x, y) = 3 \log x + \log y$, el consumidor midiese su utilidad mediante una función monótona de ésta, por ejemplo,

$g(x, y) = \exp f(x, y) = x^3 y$, las correspondientes funciones de demanda serían las mismas, como se comprueba inmediatamente en este ejemplo y se demostrará en general en el siguiente capítulo. Tratándose de funciones *funcionalmente dependientes*, las curvas de nivel de f y g son las mismas, lo cual parece indicar —como precisaremos— que son realmente éstas, y no la función de utilidad en sí misma, las que intervienen en los cálculos. De momento, basta observar que, como es típico en los problemas de optimización con restricciones de igualdad, en el óptimo (x^*, y^*) la curva de nivel $f(x, y) = \text{cte}$ es tangente a la recta presupuestaria $p_x x + p_y y = \text{cte}$ correspondiente.

EJEMPLO 3. Combinando argumentos de producción y consumo, supongamos ahora que y representa el número de horas diarias que el consumidor del ejemplo anterior dedica al ocio, que x son sus ingresos diarios y que éstos proceden de la venta, al precio de mercado p , de un producto fabricado por el propio consumidor según la función de producción $f(l) = 10l^\alpha$, siendo $l = 24 - y$ las horas de trabajo dedicadas cada día a la fabricación del producto y $\alpha \in (0, 1)$ una constante. ¿Qué elección de (x, y) le proporcionará la máxima utilidad?

En este caso no parece pertinente suponer la existencia de un “precio de mercado” para el ocio, pero se puede razonar así: la utilidad de la combinación de “bienes” $x = pf(l) = 10pl^\alpha$, $y = 24 - l$, expresada en función de l , vendrá dada por

$$U(l) = 3 \log(10pl^\alpha) + \log(24 - l) = 3 \log 10p + 3\alpha \log l + \log(24 - l)$$

que alcanza un máximo global para $l^* = \frac{72\alpha}{3\alpha + 1}$. Si, por ejemplo, $\alpha = \frac{1}{6}$, se obtiene $l^* = 8$, que representa el número de horas de trabajo que proporcionan la máxima utilidad a nuestro productor-consumidor, es decir, su combinación “ideal” entre ocio e ingreso.

Intercambio puro

La cuestión del *intercambio* entre agentes económicos es bastante más compleja que las anteriores; incluso los ejemplos numéricos más sencillos dan lugar a cálculos prohibitivos. Por ello, nos limitaremos a unos pocos comentarios ilustrativos de uno de los aspectos más interesantes (y estéticos) de la teoría marginalista: la *caja de Edgeworth*.

EJEMPLO 4. Sean C_1, C_2 dos consumidores de dos bienes x e y con funciones de utilidad respectivas $u_1 = f_1(x, y) = xy$, $u_2 = f_2(x, y) = xy^3$. Supongamos que inicialmente C_1 posee X unidades de x y ninguna de y , mientras que C_2 no posee nada de x e Y unidades de y . ¿Cómo deberán repartirse estos

consumidores la cesta total (X, Y) de forma que la utilidad de ambos sea óptima?

Fijemos de momento la utilidad U_1 del consumidor C_1 . El consumidor C_2 intentará maximizar su utilidad tomando este dato como restricción, exactamente igual que si fuese una restricción presupuestaria. Si (x, y) representa una combinación dada de bienes poseídos por C_1 , la combinación asignada a C_2 será $(X - x, Y - y)$. Por tanto, el problema de optimización que éste deberá resolver es el siguiente:

$$\text{Prob. } (U_1) \quad \begin{cases} \text{maximizar } f_2(X - x, Y - y) \\ \text{condicionado a } f_1(x, y) = U_1 \end{cases}$$

De la restricción $xy = U_1$ podemos despejar $y = U_1/x$, y el máximo de $(X - x)(Y - U_1/x)^3$ se alcanza para

$$x^* = x^*(U_1) = \frac{1}{Y} \left(-U_1 + \sqrt{U_1^2 + 3YU_1X} \right),$$

con su valor asociado $y^* = y^*(U_1) = U_1/x^*(U_1)$. La utilidad correspondiente es

$$U_2^* = U_2^*(U_1) = f_2(X - x^*(U_1), Y - y^*(U_1)) = (X - x^*(U_1))(Y - y^*(U_1))^3$$

La curva formada por los puntos $(x^*(U_1), y^*(U_1))$ y parametrizada por U_1 se llama *curva de contrato* entre C_1 y C_2 , y representa las combinaciones óptimas en el siguiente sentido: para cualquier combinación distinta a $(x^*(U_1), y^*(U_1))$, o bien la utilidad de C_1 es inferior a U_1 , o bien la de C_2 es inferior a $U_2^*(U_1)$. En otras palabras, al menos uno de los dos consumidores sale perdiendo en utilidad. Este tipo de optimalidad, en la cual se trata de mejorar la posición de alguno de los consumidores sin que pierda ninguno de los demás, se conoce con el nombre de *optimalidad de Pareto*, en honor al economista italiano Vilfredo Pareto, uno de los creadores de la teoría formal del intercambio. Con esta terminología, cualquier punto de la forma $(x^*(U_1), y^*(U_1))$ es un *óptimo de Pareto* del par (f_1, f_2) . También se dice que la combinación (x^*, y^*) es *eficiente en sentido de Pareto*, o *Pareto-eficiente*.

En realidad, más que los valores numéricos obtenidos, interesa aquí su *interpretación geométrica*. Obsérvese que la curva de nivel N de la función $g_2(x, y) \stackrel{\text{def}}{=} f_2(X - x, Y - y)$ es la figura simétrica respecto del punto (X, Y) de la curva de nivel N de f_2 . Lo más útil es, pues, representar simultáneamente las curvas de nivel $f_1(x, y) = \text{cte}$ junto con las $f_2(X - x, Y - y) = \text{cte}$, que, a su vez, se obtienen poniendo el origen de coordenadas en el punto (X, Y) y haciendo la simetría puntual respecto de (X, Y) . Con esto, el rectángulo de vértices $(0, 0)$, $(0, X)$, $(Y, 0)$ y (X, Y) aparece cubierto por dos familias de

curvas, unas referidas a $(0,0)$ y las otras a (X,Y) ; los puntos de tangencia entre ambos sistemas de curvas forman la *curva de contrato*, y son óptimos de Pareto del par (f_1, f_2) . Esta representación geométrica (figura 10.4) se llama *caja de Edgeworth*, por el economista británico que la introdujo en su obra "Psíquica Matemática", uno de los primeros intentos de cuantificar con rigor las preferencias psicológicas de los agentes individuales.

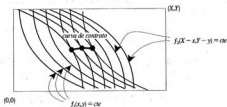


Figura 10.4 Caja de Edgeworth

Terminemos este capítulo con una reflexión: ¿Es representativo un análisis de este tipo, con una función de utilidad tan nítida? Ya desde el filósofo utilitarista Bentham (1748-1832), se discute la posibilidad de medir la "utilidad" o "satisfacción" que reporta el disfrute de determinados bienes, y como hipótesis de trabajo da lugar a una teoría muy interesante, como hemos podido atisbar (sobre todo con las ideas de Edgeworth) y veremos con detalle en el siguiente capítulo. Aun así, siempre ha sido éste un motivo de gran controversia en las ciencias del comportamiento, y es importante saber hasta qué punto el edificio entero de la teoría del consumidor reposa sobre la *existencia de una función de utilidad* que pueda realmente medirse, pues de no ser así podríamos hallarnos ante una teoría muy redonda quizá, pero inaplicable a situaciones reales.

Pues bien, la teoría de la elección racional del consumidor se puede formular de manera que la función de utilidad sólo intervenga a través de la *utilidad marginal* (cosa bastante natural, pues en las condiciones de óptimo aparecen las derivadas primeras, es decir, las magnitudes marginales asociadas), mucho más próxima a la intuición y "medible" empíricamente, al menos de forma aproximada, en muchos casos. Lo que supone la revolución marginalista en el proceso de conceptualización de la teoría económica se traduce en términos matemáticos en la sorprendente *desaparición* de la propia función de utilidad de las fórmulas que dan la elección óptima, y su sustitución por proporcionalidades entre utilidades marginales, cuestión de fundamental importancia que se analiza en profundidad en el capítulo siguiente.

Capítulo 11

Teoría de la oferta y la demanda

Toda actividad económica consiste en la producción e intercambio de bienes, ya sean éstos realmente bienes (mercancías) o servicios. En cualquier economía moderna estas actividades se realizan de manera vertiginosa y por un enorme número de agentes, por lo que una descripción exhaustiva de todas sus características estaría condenada al fracaso. Así, la primera etapa en la elaboración de un modelo teórico en Economía consiste, al igual que sucede en las demás ciencias, en hacer abstracción de la inmensa complejidad del mundo real, extrayendo los aspectos que se consideran *esenciales* para explicar el fenómeno en estudio.

En este sentido, la teoría económica divide a los agentes económicos en productores y consumidores, entendiendo por tales las unidades de toma de decisiones, respectivamente, en la producción y en el consumo. Esta clasificación no es excluyente; de hecho, la mayor parte de los productores, cuando no todos, son al mismo tiempo consumidores. Sin embargo, lo que interesa es la conducta de estas unidades de toma de decisiones: se supone que un productor (generalmente llamado empresa) está interesado en elegir la cantidad y calidad de factores (trabajo, capital y materias primas) que debe utilizar para producir un bien con el objetivo de maximizar sus beneficios; por el contrario, un consumidor (de usual un individuo) intenta seleccionar, de acuerdo con sus preferencias personales, la combinación óptima de bienes a adquirir empleando sus recursos disponibles (renta).

Estas caracterizaciones de los dos tipos de agentes económicos dejan entrever un nexo común: en ambos casos los recursos (factores de producción y renta del consumidor) son limitados y se pretende distribuirlos de manera eficiente; es decir, las diferencias entre dichos agentes no se encuentran

en la estructura lógica de su comportamiento, sino en el tipo de objetivo y en las restricciones subyacentes. La mencionada estructura formal (*elección racional*) es lo que caracteriza la conducta de cualquier agente económico; de hecho, según la definición tradicional ya comentada en el capítulo anterior, la Economía es la ciencia que se ocupa de la asignación óptima de recursos escasos entre fines alternativos.

En el presente capítulo se analiza la decisión de los dos tipos de agentes económicos: en primer lugar, la elección de los consumidores, que determina la demanda individual (por parte de un consumidor) de un bien fijado su precio y, en último lugar, la demanda (total) de mercado como agregación (suma) de las demandas individuales; finalmente, se acomete el estudio de la producción óptima, que genera la oferta del bien supuesto fijo su precio de venta.

11.1 Teoría clásica de la demanda

La teoría de la elección de los consumidores tiene su punto de partida en la modelización de su supuesta conducta racional: se considera que los individuos adquieren la combinación de bienes que más se adecúa a sus preferencias (gustos) personales. Estas preferencias se suelen formalizar por medio de una función U , denominada *función de utilidad*, la cual constituye una idealización matemática de las sensaciones subjetivas (nivel de satisfacción, placer, cese de necesidad, etc.) que se derivan del consumo. Así, $U(A) > U(B)$, esto es, que la opción A reporte más utilidad que la B , equivale a decir que A es preferida a B ; de igual forma, $U(A) = U(B)$ sugiere que el individuo es indiferente entre las opciones A y B .

Para un mercado con dos bienes, la función de tipo Cobb-Douglas

$$U(x, y) = cx^a y^b, \quad a, b, c > 0, \quad a + b < 1, \quad (11.1)$$

donde x e y denotan las cantidades respectivas consumidas de cada uno de ellos, constituye un ejemplo típico de función de utilidad. Sus curvas de nivel, llamadas *curvas de indiferencia* pues determinan las cestas de consumo (x, y) con misma utilidad, aparecen representadas en la figura 11.1. En ella se observa, en primer lugar, que las curvas de nivel que pasan por un punto (x_0, y_0) , con $x_0, y_0 > 0$, están contenidas en el cuadrante positivo del plano $\mathbb{R}_{++}^2 = \{(x, y) \in \mathbb{R}^2; x, y > 0\}$, lo que conlleva la imposibilidad de que una cesta de consumo positivo en ambos bienes sea indiferente con otra de consumo negativo o nulo en algún bien.

Además, las curvas de indiferencia $U(x, y) = u$ definen implícitamente en \mathbb{R}_{++}^2 funciones de la forma $y = f_u(x)$ (ó $x = g_u(y)$), decrecientes y estrictamente convexas. El decrecimiento de estas funciones implica que, si se desea mantener el mismo nivel de utilidad, un aumento en el consumo de un bien

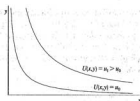


Figura 11.1 Curvas de indiferencia de la utilidad Cobb-Douglas

produce una disminución en el relativo al otro. La pendiente de las curvas de indiferencia, dada por

$$f'_u(x) = -\frac{U_x(x, y)}{U_y(x, y)} = -\frac{ay}{bx},$$

representa dicha relación de intercambio. En la terminología económica, si (x_0, y_0) es un punto de la curva $U(x, y) = u_0$, el cociente

$$RMS(x_0, y_0) \stackrel{\text{def}}{=} \frac{U_x(x_0, y_0)}{U_y(x_0, y_0)} \quad (11.2)$$

se denomina *relación marginal de sustitución*, o de manera equivalente, $RMS(x_0, y_0) = -f'_{u_0}(x_0)$. Al igual que sucede con la mayor parte de las aplicaciones del cálculo diferencial a la Economía (recuérdense las observaciones realizadas al respecto en el capítulo precedente), dicho cociente se interpreta en términos finitos: dado que

$$RMS \approx \frac{\frac{\Delta U}{\Delta x}}{\frac{\Delta U}{\Delta y}} = \frac{\Delta y}{\Delta x},$$

la relación marginal de sustitución representa las unidades que se ha de incrementar el consumo del segundo bien (Δy) para que una disminución de una unidad en el primer bien ($\Delta x = -1$) no haga variar el nivel de utilidad. Según esto, el decrecimiento de las curvas de indiferencia se traduce en una $RMS > 0$, lo que se tiene garantizado (como sucede para la utilidad de tipo Cobb-Douglas) cuando las utilidades marginales U_x y U_y son siempre positivas.

Finalmente, la convexidad estricta de las funciones implícitas $y = f_u(x)$ tiene dos importantes consecuencias. Por un lado, manifiesta preferencia por

cestas de consumo intermedias (más equilibradas); en efecto, dados dos vectores de consumo indiferentes (\hat{x}, \hat{y}) , (\bar{x}, \bar{y}) , con $U(\hat{x}, \hat{y}) = U(\bar{x}, \bar{y}) = \bar{u}$, la convexidad del conjunto $\{(x, y) \in \mathbb{R}^2; U(x, y) \geq \bar{u}\}$ implica, para cualquier consumo $(x^*, y^*) = \lambda \cdot (\hat{x}, \hat{y}) + (1 - \lambda) \cdot (\bar{x}, \bar{y})$ con $\lambda \in (0, 1)$, que $U(x^*, y^*) > U(\hat{x}, \hat{y}) = U(\bar{x}, \bar{y}) = \bar{u}$. Por otro, se traduce en una relación marginal de sustitución decreciente:

$$\frac{dRMS(x, f_u(x))}{dx} = -f_u''(x)$$

y $f_u''(x) \geq 0$ para todo $x > 0$ al ser $f_u(x)$ convexa y de clase C^2 en \mathbb{R}_{++}^2 ; así pues, la relación de intercambio entre los dos bienes es decreciente respecto al consumo.

En un contexto más general, considerando un mercado con n bienes disponibles, los anteriores comentarios quedan enunciados de la siguiente forma:

- Un vector ("cesta") típico de consumo se denota por $x = (x_1, x_2, \dots, x_n)$, donde $x_i \geq 0$ representa la cantidad consumida del bien i .
- El espacio de bienes $X \subset \mathbb{R}_+^n = \{x = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n; x_i \geq 0\}$, está formado por todos los vectores de consumo factibles; por simplicidad, en adelante consideraremos que $X = \mathbb{R}_+^n$.
- La función de utilidad $U : X \rightarrow \mathbb{R}_+$, que asocia a cada cesta x de consumo la utilidad $U(x) \geq 0$ que reporta, se supone que cumple las siguientes hipótesis:

(H.1) U es de clase C^2 sobre $\mathbb{R}_{++}^n = \{x = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n; x_i > 0\}$ y sus conjuntos de nivel que pasan por un punto de \mathbb{R}_{++}^n están totalmente contenidos en \mathbb{R}_{++}^n .

(H.2) $\frac{\partial U}{\partial x_i}(x) > 0$ para todo $x \in \mathbb{R}_{++}^n$ y cada $i = 1, \dots, n$, es decir, la función $U(x_1, x_2, \dots, x_n)$ es creciente en cada variable x_i , para $i = 1, \dots, n$. Esta hipótesis refleja la idea de que el consumidor prefiere recibir cuantos más bienes mejor (por lo que se suele denominar *hipótesis de no saciedad*) y tiene una importante implicación sobre los conjuntos de nivel de la función de utilidad, denominados *conjuntos* (curvas si $n = 2$) *de indiferencia*. En concreto, los conjuntos de indiferencia $U(x_1, x_2, \dots, x_n) = c$ definen en \mathbb{R}_{++}^n a cada variable, por ejemplo a x_n , como función implícita de las restantes ($x_n = g(x_1, x_2, \dots, x_{n-1})$), siendo $g(x_1, x_2, \dots, x_{n-1})$ una función decreciente en cada variable x_i , $i = 1, \dots, n-1$.

- (H.3) La matriz hessiana $D^2U(x)$ es definida negativa para todo $x \in \mathbb{R}_{++}^n$, esto es, U es una función estrictamente cóncava sobre \mathbb{R}_{++}^n , lo cual implica que los conjuntos $\Lambda_\alpha = \{x \in \mathbb{R}_{++}^n ; U(x) \geq \alpha\}$ son convexos para cualquier $\alpha \geq 0$. Esto quiere decir que si consideramos dos cestas de consumo indiferentes entre sí, $U(x^1) = U(x^2) = \alpha$, las cestas intermedias (más equilibradas) $x = \lambda x^1 + (1 - \lambda)x^2$, $\lambda \in (0, 1)$, pertenecen a Λ_α , por lo que reportan más utilidad. Además, como hemos comentado anteriormente, esta hipótesis también implica que las funciones implícitas generadas por los conjuntos de indiferencia son estrictamente convexas.

Estas hipótesis sobre la función de utilidad intentan formalizar matemáticamente el comportamiento del consumidor, que, en términos económicos, está fundamentado en dos principios: se supone que los consumidores son insaciables y que prefieren cestas de consumo equilibradas. Naturalmente, para el caso de un mercado con dos bienes (que tiene especial interés pues, aparte de poder representarse en el plano, con los lógicos beneficios interpretativos que esto supone, puede utilizarse para analizar la influencia, en ambas direcciones, entre la cantidad consumida de un bien x y la correspondiente al consumo agregado del resto de bienes del mercado y), la función de utilidad de tipo Cobb-Douglas cumple las hipótesis (H.1), (H.2) y (H.3). Sin embargo, estas hipótesis pueden debilitarse, eso sí, manteniendo la validez de los principios económicos, de muy diferentes formas: una de ellas consiste en requerir simplemente que la función de utilidad sea continua, no decreciente y cóncava. Como ejemplo de una función de utilidad no estrictamente cóncava podemos considerar

$$U(x, y) = ax + by, \quad a, b > 0 \quad (11.3)$$

El carácter lineal de estas curvas de indiferencia (mostradas en la figura 11.2) dio lugar al término *sustitutivos perfectos* para describir la relación entre ambos bienes. Un individuo que tuviera estas preferencias estaría dispuesto a renunciar siempre a la misma cantidad del bien x para aumentar una unidad su consumo de y . Cuando $a = b$, esta situación podría describir la relación entre diferentes marcas de idéntico producto; como muestra, nótese que la mayoría de los conductores de nuestro país son indiferentes entre adquirir 1 litro de gasolina Petronor o BP.

Las curvas de indiferencia en forma de L de la figura 11.3, generadas por la función de utilidad

$$U(x, y) = \min(ax, by), \quad a, b > 0, \quad (11.4)$$

que no es diferenciable ni (estrictamente) creciente, muestran el caso opuesto al de los sustitutivos perfectos. Estas preferencias se aplicarían a bienes indis-

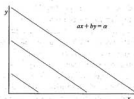


Figura 11.2 Curvas de indiferencia de bienes sustitutos perfectos

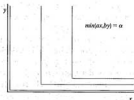


Figura 11.3 Curvas de indiferencia de bienes complementarios perfectos

luciblemente unidos; por ejemplo, si una persona tiene 10 tornillos de pared y 10 tacos, no incrementaría su utilidad el disponer de un tornillo adicional o de un taco más. Otros clásicos ejemplos de bienes *complementarios perfectos* son el zapato derecho e izquierdo y, para bastantes consumidores, el café y la leche. En esta situación, no habrá exceso de ninguno de los dos bienes únicamente cuando $ax = by$.

Como se ha comentado anteriormente, todavía se puede ir más lejos en el debilitamiento de las hipótesis de la función de utilidad. En concreto, la hipótesis (H.3) puede ser expresada en términos de funciones cuasicóncavas (véase [9]), que son, de hecho, aquellas que garantizan *sine qua non* la convexidad de los conjuntos Λ_α . Incluso, es posible generalizar el propio concepto de función de utilidad expresando las preferencias de los consumidores por medio una relación definida en el espacio de bienes (para una aproximación a esta axiomática puede consultarse [7]). En cualquier caso, con el objetivo de no desviarnos del propósito de esta parte tercera del libro, que no es otro que mostrar las aplicaciones del cálculo diferencial a la Economía, nosotros abordaremos la teoría del consumidor enmarcada en el contexto del cumplimiento

de las hipótesis (H.1), (H.2) y (H.3).

El concepto de función de utilidad es la noción fundamental de la teoría de la elección del consumidor pues permite modelizar su comportamiento formulándolo en términos de un problema de optimización: el consumidor queda reducido a un agente maximizador de utilidad, condicionado únicamente por su limitación monetaria. Esta restricción presupuestaria viene dada por

$$G(p, x) \equiv p \cdot x = \sum_{i=1}^n p_i x_i \leq r \quad (11.5)$$

donde $p_i > 0$ denota el precio del bien i (consumido en una cantidad x_i), $p = (p_1, \dots, p_n)$ es el vector de precios, $G(p, x)$ la función de gasto y $r > 0$ la cantidad total de dinero (renta) disponible. Así pues, se trata de estudiar el problema

$$\begin{cases} \text{MAX.} & U(x) \\ \text{sujeito a} & G(p, x) \leq r \\ & x \geq 0 \end{cases} \quad (11.6)$$

conocidos los valores de p y r .

Antes de abordar la resolución de este problema es conveniente realizar algunas consideraciones preliminares. En primer lugar, la existencia de solución global está garantizada en virtud del Teorema de Weierstrass: el conjunto factible

$$B = \{x \in \mathbb{R}^n; G(p, x) \leq r, x \geq 0\}$$

es cerrado y acotado ($0 \leq x_i \leq \frac{r}{p_i}$ para cada $i = 1, \dots, n$) y la función objetivo U es continua en B (hipótesis (H.1)). Además, de la convexidad de B y la concavidad estricta de U (hipótesis (H.3)) se deduce que el programa es estrictamente convezo, por lo que la solución global es única y no existen soluciones locales no globales.

Por otro lado, la hipótesis (H.2) implica que la función de utilidad no tiene puntos críticos en \mathbb{R}_{++}^n y, en consecuencia, se deduce que la solución del programa (11.6) no está en el interior del conjunto factible B . Tampoco se encuentra en la parte de la frontera de B donde $x_i = 0$ para algún $i = 1, \dots, n$, pues la condición sobre los conjuntos de indiferencia dada en (H.1) conlleva un mayor nivel de utilidad sobre \mathbb{R}_{++}^n . Así pues, se concluye que la solución global de (11.6) verifica $G(p, x) = r$.

La maximización de la utilidad

Teniendo en cuenta los comentarios anteriores, bajo las hipótesis (H.1), (H.2) y (H.3) sobre la función de utilidad, el programa (11.6) tiene única solución, es

positiva y se alcanza en un punto de saturación de la restricción presupuestaria. Así, el problema de la elección del consumidor puede ser matemáticamente formulado por el programa

$$\begin{cases} \text{MAX.} & U(x) \\ \text{sujeto a} & G(p, x) = r \end{cases} \quad (11.7)$$

más simple que el inicialmente considerado, pero con única e idéntica solución que (11.6).

NOTA. En principio se podría pensar que las hipótesis consideradas afectan irremediablemente a la formulación del problema económico: el hecho de que la solución del problema (11.6) se alcance en un punto $x^* \in \mathbb{R}_{++}^n$ tal que $G(p, x^*) = r$ parece indicar que no se está considerando la utilidad del ahorro, es decir, sugiere que el problema, tal y como está planteado, atribuye utilidad cero a cualquier cantidad de dinero no invertida en la adquisición de bienes. Sin embargo, el supuesto de gastar toda la renta disponible no es restrictivo ya que el ahorro puede ser considerado de forma totalmente correcta como uno de dichos bienes, como también sucede, por ejemplo, con el tiempo de ocio.

Los puntos estacionarios del programa (11.7) se obtienen aplicando las condiciones necesarias de primer orden de Lagrange. En concreto, si consideramos la función Lagrangiana

$$L(x, \lambda) = U(x) - \lambda \cdot (G(p, x) - r)$$

dichos puntos $(x^*, \lambda^*) \in \mathbb{R}^{n+1}$ verifican el sistema

$$\begin{cases} \frac{\partial L}{\partial x_i}(x, \lambda) = \frac{\partial U}{\partial x_i}(x) - \lambda p_i = 0 & (i = 1, \dots, n) \\ \frac{\partial L}{\partial \lambda}(x, \lambda) = -G(p, x) + r = 0 \end{cases} \quad (11.8)$$

de donde se deduce que

$$\frac{\frac{\partial U}{\partial x_i}(x^*)}{p_i} = \frac{\frac{\partial U}{\partial x_j}(x^*)}{p_j} \quad (11.9)$$

para cada par de bienes distintos $i \neq j$, siendo el multiplicador de Lagrange asociado

$$\lambda^* = \frac{\frac{\partial U}{\partial x_i}(x^*)}{p_i} > 0 \quad (11.10)$$

Para el caso de dos bienes ($n = 2$), el problema se escribe

$$\begin{cases} \text{MAX.} & U(x, y) \\ \text{sujeto a} & p_1 x + p_2 y = r \end{cases} \quad (11.11)$$

y la condición (11.9)

$$\frac{\frac{\partial U}{\partial x}(x^*, y^*)}{\frac{\partial U}{\partial y}(x^*, y^*)} = \frac{p_1}{p_2} \quad (11.12)$$

equivale a decir que, en el óptimo, la pendiente de la curva de nivel de la función de utilidad (dada por $-\frac{\frac{\partial U}{\partial x}(x^*, y^*)}{\frac{\partial U}{\partial y}(x^*, y^*)}$) coincide con la pendiente de la recta presupuestaria ($-\frac{p_1}{p_2}$), es decir, en ese punto la curva de indiferencia es tangente a la recta de gasto (véase la figura 11.4). En términos económicos, la relación marginal de sustitución debe ser igual a la relación de precios de los bienes.

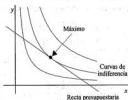


Figura 11.4 Maximización de la utilidad en un mercado con dos bienes

Volviendo al caso general de n bienes, nótese que se cumple la condición de regularidad en todo el conjunto factible ($D_x G(p, x) = p \neq 0$), por lo que el máximo global (cuya existencia está garantizada) se alcanza en un punto estacionario del programa; además, la concavidad estricta de la función de utilidad implica unicidad de óptimo (local y global) y, en conjunción con lo anterior, existencia de un único punto estacionario. Por tanto, el sistema (11.8) tiene definida única solución para cada valor de los parámetros $p > 0$ y $r > 0$, en la que se alcanza el máximo del programa (11.7). Además, la aplicación del Teorema de la función implícita a dicho sistema,

$$D_{x,\lambda} L(x, \lambda, p, r) = 0$$

cuya matriz jacobiana

$$D_{x,\lambda}^2 L(x, \lambda, p, r) = \begin{pmatrix} D^2 U(x) & -p \\ -p^t & 0 \end{pmatrix} \quad (11.13)$$

tiene determinante no nulo, permite expresar la solución como una función C^1 de los parámetros p y r (para demostrar que esta matriz "particionada

por cajas", donde se entiende que p es una matriz columna, es invertible se pueden utilizar las propiedades de las matrices particionadas; consúltase para este punto [2]).

En consecuencia, el problema de la maximización de la utilidad genera las siguientes funciones (todas ellas, al menos, de clase C^1): la función solución $x^M(p, r) : \mathbb{R}_{++}^{n+1} \rightarrow \mathbb{R}^n$, llamada *función marshalliana de demanda*, que determina lo que el consumidor adquirirá y consumirá fijados unos precios p y una renta r ; la función valor óptimo $U^M(p, r) = U[x^M(p, r)] : \mathbb{R}_{++}^{n+1} \rightarrow \mathbb{R}$, denominada *función de utilidad indirecta*, que indica el nivel de utilidad así alcanzado; y la *función multiplicador* $\lambda^M(p, r) : \mathbb{R}_{++}^{n+1} \rightarrow \mathbb{R}$, dada por (11.10).

EjemPlo. Consideremos la función de utilidad Cobb-Douglas $U(x, y) = 3x^{\frac{1}{3}}y^{\frac{2}{3}}$; en este caso el sistema (11.8) se escribe como

$$\begin{cases} x^{-\frac{2}{3}}y^{\frac{2}{3}} - p_1\lambda = 0 \\ 2x^{\frac{1}{3}}y^{-\frac{1}{3}} - p_2\lambda = 0 \\ p_1x + p_2y = r \end{cases}$$

de donde se deduce que

$$\lambda = \frac{x^{-\frac{2}{3}}y^{\frac{2}{3}}}{p_1} = \frac{2x^{\frac{1}{3}}y^{-\frac{1}{3}}}{p_2} \implies p_2y = 2p_1x$$

y sustituyendo en la última ecuación resulta

$$x^M(p, r) = \frac{r}{3p_1} \quad ; \quad y^M(p, r) = \frac{2r}{3p_2}$$

siendo

$$\begin{aligned} \lambda^M(p, r) &= \frac{(x^M)^{-\frac{2}{3}}(y^M)^{\frac{2}{3}}}{p_1} = \frac{2^{\frac{2}{3}}}{p_1^{\frac{1}{3}}p_2^{\frac{2}{3}}} \\ U^M(p, r) &= U(x^M, y^M) = \frac{2^{\frac{2}{3}}r}{p_1^{\frac{1}{3}}p_2^{\frac{2}{3}}} \end{aligned}$$

Obsérvese el hecho curioso (y atípico) de que la demanda de cada bien depende exclusivamente de su propio precio (y no del precio del otro).

Aunque, por ejemplo, para las funciones de utilidad de tipo Cobb-Douglas es posible obtener explícitamente las funciones de demanda y de utilidad indirecta, la verdadera importancia de esta teoría del consumidor descansa en la posibilidad de efectuar estos cálculos en general, gracias a las hipótesis realizadas, sin necesidad de conocer explícitamente la forma de $U(x)$. Además, las funciones generadas a partir del problema de la maximización de la utilidad tienen ciertas propiedades que merecen tenerse en cuenta:

- (i) La función marshalliana de demanda es homogénea de grado cero. En efecto, $x = x^M(p, r)$ se corresponde con la única solución del sistema con n ecuaciones y n incógnitas

$$\begin{cases} \frac{\partial U}{\partial x_i}(x) = \frac{p_i}{p_j} & i, j = 1, \dots, n \ (i \neq j) \\ p \cdot x = r \end{cases}$$

que coincide con la solución del sistema

$$\begin{cases} \frac{\partial U}{\partial x_i}(x) = \frac{tp_i}{tp_j} & i, j = 1, \dots, n \ (i \neq j) \\ (tp) \cdot x = tr \end{cases}$$

para todo $t > 0$, por lo que $x^M(tp, tr) = x^M(p, r)$. Esto significa que si se multiplican tanto los precios como la renta por una misma cantidad, el consumidor elegirá la misma cesta de bienes, considerando que no ha cambiado nada en su situación "real" y que sólo ha habido un "cambio de escala" en el dinero.

- (ii) Como consecuencia directa de lo anterior se deduce que la función de utilidad indirecta también es homogénea de grado cero, pues para cada $t > 0$ se verifica $U^M(tp, tr) = U[x^M(tp, tr)] = U[x^M(p, r)] = U^M(p, r)$.
- (iii) Utilizando las propiedades de los multiplicadores de Lagrange se obtiene una relación (denominada *identidad de Roy*) que expresa (y, por tanto, permite determinar) la función marshalliana de demanda a partir de la función de utilidad indirecta. En concreto, aplicando el teorema de la envolvente (véase la nota 2 que sigue al teorema 6.15) respecto a p_i se tiene

$$\frac{\partial U^M}{\partial p_i} = \frac{\partial L}{\partial p_i}(x^M(p, r), \lambda^M(p, r)) = -\lambda^M(p, r) \cdot x_i^M(p, r) \quad i = 1, \dots, n$$

y respecto a r

$$\frac{\partial U^M}{\partial r} = \frac{\partial L}{\partial r}(x^M(p, r), \lambda^M(p, r)) = \lambda^M(p, r)$$

de donde se concluye que

$$x_i^M(p, r) = -\frac{\frac{\partial U^M}{\partial p_i}}{\frac{\partial U^M}{\partial r}} \quad i = 1, \dots, n \quad (11.14)$$

Como la propia notación pone de manifiesto, la demanda de un producto $x_i^M(p, r)$ depende del precio de los bienes del mercado y de la renta del consumidor, por lo que cambios en los valores de estos parámetros afectan a aquella. En general, los cambios en la demanda como respuesta a variaciones de los precios son más complejos de analizar; su estudio se llevará a cabo más adelante. Sin embargo, los cambios producidos por variaciones en la renta se pueden analizar directamente a partir del problema de la maximización de la utilidad ya estudiado: cuando aumenta la renta (poder adquisitivo) de un consumidor, la restricción presupuestaria dada por el hiperplano

$$G(p, x) = \sum_{i=1}^n p_i x_i = r$$

se desplaza paralelamente, lo que produce un desplazamiento en la solución del problema (la demanda) obtenida como punto de tangencia entre los conjuntos de indiferencia y dicho hiperplano.

En principio, parece natural esperar que si aumenta la renta, también aumente la cantidad demandada del producto; esto se ilustra, para el caso de un mercado con dos bienes, en la figura 11.5. Como se observa, el desplazamiento

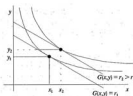


Figura 11.5

paralelo de las rectas de gasto produce un aumento en la cantidad consumida de ambos bienes.

Ahora bien, y aunque a primera vista pueda parecer sorprendente, también es posible que los incrementos en la renta provoquen disminuciones en la demanda. En la figura 11.6 se muestra esta situación: la cantidad demandada del bien x disminuye al pasar de un nivel de renta r_1 a $r_2 > r_1$.

En consecuencia, el signo de la derivada

$$\frac{\partial x_i^M(p, r)}{\partial r}$$

depende de la forma de la función de utilidad. En términos económicos, si esta derivada es negativa se dice que el bien i es un *bien inferior*; en caso

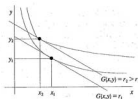


Figura 11.6

contrario, se denomina *bien normal*. Así, un bien inferior es aquel que disminuye su consumo al aumentar la renta disponible; ejemplos bien elocuentes son el whisky de garrafa y el sucedáneo de caviar. En general, el calificativo de inferior o normal para un bien depende de los niveles considerados para la renta: un bien puede ser inferior para ciertos valores de r y normal para otros; como muestra, nótese que las hamburguesas representan posiblemente un bien inferior para un norteamericano medio y un bien normal para la mayor parte de la población africana.

El estudio de la demanda de un producto en función de la renta del consumidor se puede llevar todavía más lejos. Fijados los precios de los bienes del mercado, la función $x_i^M(r)$ genera una curva en el plano conocida como *curva de Engel*. El crecimiento-decrecimiento de esta curva determina el calificativo de normal-inferior para un bien. Su concavidad-convexidad proporciona otro tipo de información: si

$$\frac{\partial^2 x_i^M(p, r)}{\partial r^2} < 0,$$

entonces la fracción de la renta destinada a la compra del bien i disminuye al aumentar la renta, lo que caracteriza a un *bien necesario*. El caso contrario pone de manifiesto el carácter de *bien de lujo*. En la figura 11.7 se muestra la curva de Engel de un bien normal, que para $r < r_0$ es un bien de lujo y para $r > r_0$ necesario. Como se observa, también en este caso el calificativo de necesario o de lujo para un bien depende del nivel de renta considerado: por ejemplo, un televisor es un bien de lujo para consumidores con rentas bajas y necesario para rentas medias y altas.

Engel realizó el primer estudio empírico riguroso sobre la conducta de los consumidores, donde observó que, en individuos con rentas no muy bajas, el porcentaje de renta dedicado a alimentación disminuye al aumentar la renta. Esta hipótesis, que pone de manifiesto que, para consumidores con renta por encima de un cierto umbral, la alimentación puede ser considerada como un

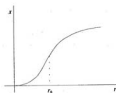


Figura 11.7

bien agregado (normal) necesario y que es conocida como *ley de Engel*, se ha contrastado en cientos de estudios. De hecho, parece tan plausible que algunos economistas han sugerido que la proporción de la renta destinada a alimentación podría ser un efectivo indicador de la pobreza: si esta proporción es mayor que una cierta cantidad (se suele barajar la cifra del 35 o 40 por ciento), el consumidor (o el país, si se trata de un consumidor medio) debería ser considerado 'pobre'.

La minimización del gasto

El problema de la elección del consumidor, que ha sido expresado en términos de maximización de la utilidad, admite una formulación equivalente (*dual* en la terminología matemática). Para ello, consideramos el problema de minimizar el gasto supuesto un nivel mínimo de utilidad prefijado, que se escribe de la forma

$$\begin{cases} \text{MIN.} & G(p, x) \\ \text{sujeto a} & U(x) \geq u \\ & x \geq 0 \end{cases} \quad (11.15)$$

con $p > 0$ y $u > 0$ dados. En principio, parece claro que los problemas (11.6) y (11.15) están estrechamente relacionados: si se fija como valor mínimo de la utilidad $u = U^M(p, r)$, la solución de (11.15) ha de coincidir con la de (11.6), pues de lo contrario no se estaría maximizando la utilidad. En esta sección analizaremos las relaciones entre los problemas de maximización de la utilidad y minimización del gasto, así como las soluciones y propiedades de este último.

Comencemos abordando la resolución del programa (11.15). En primer lugar, supondremos que $u > U(0)$; nótese que si $u \leq U(0)$ entonces el problema es equivalente al que se obtiene sustituyendo la restricción $U(x) \geq u$ por $U(x) \geq U(0)$ y, en este caso, la solución es trivial ($x = 0$). En segundo lugar, las hipótesis (H.1)–(H.3) permiten una simplificación en (11.15) similar a la efectuada en el problema de la maximización de la utilidad: la estricta

convexidad del conjunto factible y la acotación inferior de la función de gasto sobre dicho conjunto cerrado garantizan existencia de un único mínimo global; además, como el conjunto $\{x \in \mathbb{R}^n; U(x) \geq u\}$ está contenido en \mathbb{R}_{++}^n y no existen soluciones interiores, el problema (11.15) puede escribirse de la forma

$$\begin{cases} \text{MIN.} & G(p, x) = p \cdot x \\ \text{sujeto a} & U(x) = u \end{cases} \quad (11.16)$$

Considerando la función Lagrangiana

$$\bar{L}(x, \lambda, p, u) = px - \lambda \cdot (U(x) - u)$$

se tiene que la solución de (11.16) se corresponde con la única solución (x^{**}, λ^{**}) del sistema

$$\begin{cases} \frac{\partial \bar{L}}{\partial x_i} = p_i - \lambda \cdot \frac{\partial U}{\partial x_i}(x) = 0 & (i = 1, \dots, n) \\ \frac{\partial \bar{L}}{\partial \lambda} = -U(x) + u = 0 \end{cases} \quad (11.17)$$

que, de nuevo, satisface

$$\frac{\frac{\partial U}{\partial x_i}(x^{**})}{p_i} = \frac{\frac{\partial U}{\partial x_j}(x^{**})}{p_j} \quad (11.18)$$

para cada par de bienes distintos $i \neq j$, siendo ahora el multiplicador de Lagrange asociado

$$\lambda^{**} = \frac{p_i}{\frac{\partial U}{\partial x_i}(x^{**})} > 0 \quad (11.19)$$

Para el caso de dos bienes ($n = 2$), el problema se escribe

$$\begin{cases} \text{MIN.} & p_1x + p_2y \\ \text{sujeto a} & U(x, y) = u \end{cases} \quad (11.20)$$

y la condición (11.18) tiene la misma interpretación que en el problema de la maximización de la utilidad: en el óptimo, la recta de gasto es tangente a la curva de indiferencia (Figura 11.8). Nótese las similitudes y diferencias entre ambas situaciones (ahora la utilidad es fija y se mueven las rectas presupuestarias).

La aplicación del teorema de la función implícita al sistema (11.17) permite generar las siguientes funciones (todas ellas, al menos, de clase C^1): la función solución $x^H(p, u) : \mathbb{R}_{++}^{n+1} \rightarrow \mathbb{R}^n$, denominada *función hicksiana* de

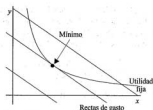


Figura 11.8 Minimización del gasto en un mercado con dos bienes

demanda —en honor de John Richard Hicks, economista inglés Premio Nobel en 1972, autor de la mayor parte de los desarrollos formales de la teoría del consumidor—, que determina lo que el consumidor adquirirá y consumirá fijados unos precios p y una utilidad (mínima) $u > U(0)$; la función valor óptimo $G^H(p, u) = G[x^H(p, u)] : \mathbb{R}_{++}^{n+1} \rightarrow \mathbb{R}$, denominada *función de gasto del consumidor*, que indica el nivel de gasto así alcanzado; y la *función multiplicador* $\lambda^H(p, u) : \mathbb{R}_{++}^{n+1} \rightarrow \mathbb{R}$, dada por (11.19).

EJEMPLO. Consideremos la función de utilidad Cobb-Douglas $U(x, y) = 3x^{\frac{1}{3}}y^{\frac{2}{3}}$; en este caso el sistema (11.17) se escribe como

$$\begin{cases} p_1 - \lambda x^{-\frac{2}{3}}y^{\frac{2}{3}} = 0 \\ p_2 - 2\lambda x^{\frac{1}{3}}y^{-\frac{1}{3}} = 0 \\ 3x^{\frac{1}{3}}y^{\frac{2}{3}} = u \end{cases}$$

de donde se deduce que

$$\lambda = \frac{p_1}{x^{-\frac{2}{3}}y^{\frac{2}{3}}} = \frac{p_2}{2x^{\frac{1}{3}}y^{-\frac{1}{3}}} \implies p_2y = 2p_1x$$

y sustituyendo en la última ecuación resulta

$$x^H(p, u) = \frac{u \cdot p_2^{\frac{2}{3}}}{2^{\frac{2}{3}}p_1^{\frac{1}{3}}} \quad ; \quad y^H(p, u) = \frac{u \cdot p_1^{\frac{1}{3}}}{2^{\frac{1}{3}}p_2^{\frac{2}{3}}}$$

siendo

$$\begin{aligned} \lambda^H(p, u) &= \frac{p_1}{(x^H)^{-\frac{2}{3}}(y^H)^{\frac{2}{3}}} = 2^{\frac{2}{3}}p_1^{\frac{1}{3}}p_2^{\frac{2}{3}} \\ G^H(p, u) &= G(x^H, y^H) = \frac{3p_1^{\frac{1}{3}}p_2^{\frac{2}{3}}u}{2^{\frac{1}{3}}} \end{aligned}$$

Las propiedades de las funciones generadas a partir del problema de la minimización del gasto se enuncian a continuación:

- (i) La función de demanda hicksiana $x^H(p, u)$ es homogénea de grado cero en la variable p , para cada $u > U(0)$.
- (ii) La función de gasto $G^H(p, u)$ es homogénea de grado uno en la variable p , para cada $u > U(0)$.
- (iii) $x_i^H(p, u) = \frac{\partial G^H}{\partial p_i}$ para cada $i = 1, \dots, n$ y para todo $u > U(0)$.
- (iv) La función de gasto $G^H(p, u)$ es cóncava en la variable p , para cada $u > U(0)$.

Las dos primeras propiedades se deducen de manera inmediata. La tercera, conocida como *Lema de Shephard para los consumidores*, es una aplicación directa del teorema de la envolvente que permite determinar la demanda hicksiana a partir de la función de gasto, y simultáneamente concluir que la función de gasto es de clase C^2 . La última, que como veremos después tiene importantes consecuencias, se demuestra a continuación: dado

$$p = \lambda p^1 + (1 - \lambda)p^2, \quad 0 \leq \lambda \leq 1$$

se tiene que

$$G^H(p, u) = p \cdot x^H(p, u) = \lambda p^1 \cdot x^H(p, u) + (1 - \lambda)p^2 \cdot x^H(p, u),$$

siendo $p^i \cdot x^H(p, u)$ el coste de adquirir $x^H(p, u)$ a precios p^i ; como $G^H(p^i, u)$ determina el coste mínimo de comprar bienes (fijada una utilidad mínima u) a precios p^i , se deduce que, naturalmente,

$$p^i \cdot x^H(p, u) \geq G^H(p^i, u), \quad i = 1, 2,$$

de donde resulta

$$\begin{aligned} G^H(p, u) &= \lambda p^1 \cdot x^H(p, u) + (1 - \lambda)p^2 \cdot x^H(p, u) \geq \\ &\geq \lambda G^H(p^1, u) + (1 - \lambda)G^H(p^2, u), \end{aligned}$$

obteniéndose que $G^H(p, u)$ es cóncava en la variable p .

La concavidad y regularidad (C^2) de la función de gasto $G^H(p, u)$ conllevan dos cualidades de su matriz hessiana: es simétrica y semidefinida negativa. Aplicando el lema de Shephard, esta matriz $D_p^2 G^H(p, u)$, denominada *matriz de Slutsky*, viene dada por

$$S(p, u) = \begin{pmatrix} \frac{\partial x_1^H}{\partial p_1} & \dots & \frac{\partial x_1^H}{\partial p_n} \\ \vdots & & \vdots \\ \frac{\partial x_n^H}{\partial p_1} & \dots & \frac{\partial x_n^H}{\partial p_n} \end{pmatrix}$$

y, por tanto, se verifica que los elementos de su diagonal principal no pueden ser positivos, es decir,

$$\frac{\partial x_i^H}{\partial p_i} \leq 0 \quad \text{para cada } i = 1, \dots, n \quad (11.21)$$

(nótese que si la matriz asociada a una forma cuadrática tiene un elemento positivo en la diagonal principal, que reordenando las variables puede situarse en la primera posición, entonces el primer menor principal sería positivo, lo que no es posible si la forma cuadrática es semidefinida negativa; complétense los detalles como ejercicio). Así pues, el incremento del precio de un bien produce necesariamente una disminución en su demanda hicksiana.

En contraposición, el signo de los restantes elementos de la matriz de Slutsky

$$\frac{\partial x_i^H}{\partial p_j} = \frac{\partial x_j^H}{\partial p_i}, \quad i \neq j$$

no está caracterizado de manera general; de hecho, conduce a una conocida clasificación de tipos de bienes. Si para dos bienes i, j (con $i \neq j$) se tiene que

$$\frac{\partial x_i^H}{\partial p_j} = \frac{\partial x_j^H}{\partial p_i} > 0$$

se dice que los bienes son *sustitutivos hicksianos netos*; en este caso, el aumento del precio de un bien produce un aumento en la demanda del otro, lo que pone de manifiesto, en cierto sentido, el carácter sustitutivo que tiene un bien respecto del otro, como cabría esperar en el manido ejemplo de la mantequilla y la margarina. Por el contrario, si

$$\frac{\partial x_i^H}{\partial p_j} = \frac{\partial x_j^H}{\partial p_i} < 0$$

se dice que los bienes son *complementarios hicksianos netos*; ahora, el aumento del precio de un bien produce una disminución tanto en su demanda como en la demanda del otro bien, lo que sucede cuando las demandas de ambos bienes están estrecha y directamente relacionadas.

Para finalizar, vamos a analizar las relaciones existentes entre los problemas de maximización de la utilidad y minimización del gasto. Estos problemas representan expresiones equivalentes (duales) de la teoría de la elección del consumidor, por lo que sus soluciones deben ser, en cierto sentido, análogas. En concreto, basta una detenida inspección a sus respectivas formulaciones para comprobar (se deja como ejercicio instructivo para el lector) las siguientes relaciones de dualidad:

$$(D.1) \quad x_i^M(p, G^H(p, u)) = x_i^H(p, u) \quad \text{para todo } i = 1, \dots, n.$$

$$(D.2) \quad U^M(p, G^H(p, u)) = u$$

$$(D.3) \quad x_i^H(p, U^M(p, r)) = x_i^M(p, r) \quad \text{para todo } i = 1, \dots, n.$$

$$(D.4) \quad G^H(p, U^M(p, r)) = r$$

La ecuación de Slutsky

En esta sección abordaremos una cuestión que había quedado pendiente con anterioridad, a saber, cómo afectan a la demanda marshalliana cambios en los precios de los bienes. En principio, parece razonable esperar que aumentos en el precio de un producto conducen a la disminución de su demanda. Sin embargo, el economista inglés Robert Giffen observó que cuando se produjo una subida del precio de las patatas en la Irlanda del siglo XIX, la población aumentó su consumo. Este hecho se conoce desde entonces con el nombre de '*paradoja de Giffen*', apelativo que hacía referencia a la imposibilidad de explicar este fenómeno con la Teoría Económica de la época. La explicación llegó de la mano del economista ruso Eugen Slutsky, quien a finales del siglo XIX consiguió identificar los dos efectos que producen las variaciones del precio de un bien sobre su demanda.

Para obtener esta descomposición, se parte de la relación de dualidad (D.1)

$$x_i^M(p, G^H(p, u)) = x_i^H(p, u)$$

y se deriva parcialmente respecto al precio p_j :

$$\frac{\partial x_i^M}{\partial p_j}(p, G^H(p, u)) + \frac{\partial x_i^M}{\partial r}(p, G^H(p, u)) \cdot \frac{\partial G^H}{\partial p_j}(p, u) = \frac{\partial x_i^H}{\partial p_j}(p, u)$$

Aplicando el lema de Shephard ($\frac{\partial G^H}{\partial p_j}(p, u) = x_j^H(p, u)$) queda

$$\frac{\partial x_i^M}{\partial p_j}(p, G^H(p, u)) + x_j^H(p, u) \cdot \frac{\partial x_i^M}{\partial r}(p, G^H(p, u)) = \frac{\partial x_i^H}{\partial p_j}(p, u)$$

Por (D.2) y (D.4), si escribimos $r = G^H(p, u)$, entonces $u = U^M(p, r)$. Finalmente, reordenando términos resulta

$$\frac{\partial x_i^M}{\partial p_j}(p, r) = \frac{\partial x_i^H}{\partial p_j}(p, U^M(p, r)) - x_j^H(p, U^M(p, r)) \cdot \frac{\partial x_i^M}{\partial r}(p, r) \quad (11.22)$$

Esta ecuación, denominada *ecuación de Slutsky*, pone de manifiesto que los cambios en los precios de los bienes tienen dos efectos sobre la demanda marshalliana:

- (i) El llamado *efecto sustitución*, dado por el término $\frac{\partial x_i^H}{\partial p_j}(p, U^M(p, r))$, que viene a medir el "grado de sustitución" entre los bienes i y j .
- (ii) El segundo término es conocido como el *efecto renta*. Muestra que el incremento de precios produce una pérdida de poder adquisitivo del consumidor, es decir, una disminución de su renta real. En concreto, $-\frac{\partial x_i^H}{\partial r}(p, r)$ representa la disminución de la demanda por unidad perdida de renta y $x_j^H(p, U^M(p, r))$ la bajada total de renta.

Para el caso particular de $i = j$, (11.22) queda

$$\frac{\partial x_i^M}{\partial p_i}(p, r) = \frac{\partial x_i^H}{\partial p_i}(p, U^M(p, r)) - x_i^H(p, U^M(p, r)) \cdot \frac{\partial x_i^M}{\partial r}(p, r), \quad (11.23)$$

expresión que permite extraer interesantes consecuencias. En primer lugar nótese que, por (11.21), el efecto sustitución es siempre no negativo. Sin embargo, el signo del efecto renta depende de si el bien es normal (efecto renta negativo) o inferior (efecto renta positivo). Claramente se desprende que los bienes normales (y muchos inferiores) cumplen

$$\frac{\partial x_i^M}{\partial p_i}(p, r) < 0$$

La única posibilidad para que esta derivada sea positiva es que el bien sea inferior y que el efecto renta sea (en valor absoluto) mayor que el sustitución. Este tipo de bienes inferiores, denominados *bienes Giffen*, permiten explicar la paradoja del mismo nombre: las patatas constituían un bien inferior que absorbía gran parte de la renta de la población irlandesa, por lo que el incremento de su precio provocó una seria bajada de la renta real que obligó a reducir el consumo de otros bienes (de lujo) con el objeto de comprar más patatas.

11.2 Teoría clásica de la oferta

Este apartado está dedicado a examinar la producción y la oferta de bienes económicos. Como veremos, buena parte del desarrollo de la teoría de la producción es similar al realizado para la teoría del consumidor. De hecho, las propias formulaciones iniciales tienen bastante paralelismo: el papel que en la elección del consumidor desempeñaba la función de utilidad ahora lo tiene la *función de producción*. En concreto, la función de producción constituye la representación matemática de la relación entre los factores utilizados y el producto obtenido. De nuevo, como ejemplo de función de producción podemos

considerar una de tipo Cobb-Douglas

$$f(K, L) = cK^a L^b, \quad a, b, c > 0, \quad a + b < 1,$$

donde K y L representan los factores utilizados (en este caso, K puede ser el número de horas de funcionamiento de las máquinas y L el número de horas empleadas por los trabajadores) y $f(K, L)$ determina la producción obtenida del bien fabricado. El decrecimiento y convexidad de sus curvas de nivel $K = g_q(L)$ (llamadas *isocuantas*), definidas implícitamente por $f(K, L) = q$, se interpretan ahora en términos de la *relación marginal de sustitución técnica*:

$$RST(K_0, L_0) \stackrel{\text{def}}{=} \frac{f_L(K_0, L_0)}{f_K(K_0, L_0)} = -g'_q(L_0)$$

que viene a medir el número de unidades que se ha de incrementar el capital (K) para que una disminución de una unidad en el trabajo (L) no haga variar el nivel de producción. Así pues, la función de producción Cobb-Douglas genera una RST positiva y decreciente.

En un contexto general, denotando por $x = (x_1, \dots, x_n) \in \mathbb{R}_+^n$ el vector de inputs utilizados (tanto el capital y el trabajo invertidos como las cantidades de las materias primas utilizadas), se supone que la función de producción $f(x)$ satisface las siguientes hipótesis:

- (H.1) f es continua en \mathbb{R}_+^n y de clase C^2 sobre \mathbb{R}_{++}^n , $f(0) = 0$ y sus conjuntos de nivel que pasan por un punto de \mathbb{R}_{++}^n están totalmente contenidos en \mathbb{R}_{++}^n .
- (H.2) $\frac{\partial f}{\partial x_i}(x) > 0$ para todo $x \in \mathbb{R}_{++}^n$ y cada $i = 1, \dots, n$, es decir, la función $f(x_1, x_2, \dots, x_n)$ es creciente en cada variable x_i , para $i = 1, \dots, n$. Esta hipótesis refleja la idea de que los inputs son productivos: un aumento de los mismos genera un aumento del producto final. Nótese que (H.1) y (H.2) implican que $f(x) > 0$ para $x \in \mathbb{R}_{++}^n$.
- (H.3) La matriz hessiana $D^2 f(x)$ es definida negativa para todo $x \in \mathbb{R}_{++}^n$, por lo que f es una función estrictamente cóncava sobre \mathbb{R}_{++}^n y los conjuntos $\Lambda_\alpha = \{x \in \mathbb{R}_{++}^n; f(x) \geq \alpha\}$ son convexos para cualquier $\alpha \geq 0$.

La función de producción constituye el pilar fundamental sobre el que se asienta la teoría de la empresa (agente económico productor, por simplicidad, de un único bien): ésta debe elegir la cantidad $q \geq 0$ de bien final a producir y las cantidad x de inputs a utilizar, conocidos los precios unitarios p y $\pi = (\pi_1, \dots, \pi_n)$ respectivos. Al vector $(q, x) \in \mathbb{R}_+^{n+1}$ se le denomina *plan de producción*, que exige a la empresa afrontar un gasto de $\pi \cdot x$ y permite obtener

(supuesto que todo lo producido se vende) unos ingresos $p \cdot q$. Se considera que la empresa tiene como objetivo maximizar los beneficios, que vienen dados por

$$B(p, \pi, q, x) = p \cdot q - \pi \cdot x = p \cdot q - \sum_{i=1}^n \pi_i x_i \quad (11.24)$$

Si la empresa dispone de financiación ilimitada (de nuevo, otro supuesto simplificador), la única restricción de la empresa es de tipo tecnológico:

$$q \leq f(x)$$

y el problema de decisión de la empresa puede enunciarse como

$$\begin{cases} \text{MAX.} & B(p, \pi, q, x) \\ \text{sujeto a} & q \leq f(x) \\ & q \geq 0, x \geq 0 \end{cases} \quad (11.25)$$

conocidos los valores de $p > 0$ y $\pi > 0$. Al no ser el conjunto factible acotado, la existencia de solución se convierte en un punto delicado; sin embargo, la consideración de una hipótesis adicional:

(H.4) $f(x) \leq M$ para todo $x \in \mathbb{R}_+^n$, lo que se suele entender como un supuesto de carácter tecnológico

permite garantizar dicho extremo (véase [9]).

Finalmente, por un lado, las hipótesis (H.1) y (H.2) implican que, en el óptimo, $q > 0$ y $x > 0$; por otro, como B no tiene puntos críticos en \mathbb{R}_{++}^n , la solución de este problema se debe alcanzar en la zona $q = f(x)$ (de lo contrario se estarían empleando de manera no eficiente los recursos disponibles). Así pues, el programa (11.25) puede reducirse a

$$\text{MAX. } \bar{B}(p, \pi, x) = B(p, \pi, f(x), x) = p \cdot f(x) - \pi \cdot x \quad (11.26)$$

que, bajo el supuesto anterior, tiene solución única pues \bar{B} es estrictamente cóncava al ser $D^2 \bar{B}(p, \pi, x) = p \cdot D^2 f(x)$ definida negativa.

La maximización de los beneficios

Teniendo en cuenta los comentarios anteriores, bajo las hipótesis (H.1), (H.2), (H.3) y (H.4) sobre la función de producción, el problema de la decisión de la empresa puede ser matemáticamente formulado por el programa (11.26), más simple que el inicialmente considerado, pero con única e idéntica solución que (11.25).

Los puntos críticos del programa (11.26) satisfacen

$$D\bar{B}(p, \pi, x) = p \cdot Df(x) - \pi = 0 \quad (11.27)$$

es decir,

$$\frac{\partial f(x)}{\partial x_i} = \frac{\pi_i}{p}, \quad i = 1, \dots, n \quad (11.28)$$

La aplicación del teorema de la función implícita a este sistema permite generar las siguientes funciones (todas ellas, al menos, de clase C^1): la función solución $x^*(p, \pi) : \mathbb{R}_{++}^{n+1} \rightarrow \mathbb{R}^n$, llamada *función de demanda de factores*, que determina los inputs que la empresa demandará; la *función de oferta del bien* $q^*(p, \pi) = f(x^*(p, \pi))$; y la función valor óptimo $B^*(p, \pi) = \bar{B}[p, \pi, x^*(p, \pi)] : \mathbb{R}_{++}^{n+1} \rightarrow \mathbb{R}$, denominada *función de beneficios de la empresa*.

Procediendo de manera análoga a como se hizo en el apartado anterior se pueden demostrar las siguientes propiedades de las funciones generadas a partir del problema de la maximización de beneficios:

- (i) Las funciones de demanda de factores $x^*(p, \pi)$ y de oferta del bien $q^*(p, \pi)$ son homogéneas de grado cero.
- (ii) La función de beneficios de la empresa $B^*(p, \pi)$ es homogénea de grado uno.
- (iii) Aplicando (11.27) se deduce que

$$\begin{aligned} \frac{\partial B^*}{\partial p}(p, \pi) &= \frac{\partial \bar{B}}{\partial p} + \sum_{i=1}^n \frac{\partial \bar{B}}{\partial x_i} \cdot \frac{\partial x_i^*}{\partial p} = q^*(p, \pi) \\ \frac{\partial B^*}{\partial \pi_i}(p, \pi) &= \frac{\partial \bar{B}}{\partial \pi_i} + \sum_{i=1}^n \frac{\partial \bar{B}}{\partial x_i} \cdot \frac{\partial x_i^*}{\partial \pi_i} = -x_i^*(p, \pi) \end{aligned}$$

resultado conocido como *lema de Hotelling*.

- (iv) La función de beneficios $B^*(p, \pi)$ es convexa y de clase C^2 .

Como consecuencia de (iv), la matriz hessiana

$$D^2 B^*(p, \pi) = \begin{pmatrix} \frac{\partial^2 q^*}{\partial p^2} & \frac{\partial^2 q^*}{\partial p \partial \pi_1} & \cdots & \frac{\partial^2 q^*}{\partial p \partial \pi_n} \\ -\frac{\partial x_1^*}{\partial p} & -\frac{\partial x_1^*}{\partial \pi_1} & \cdots & -\frac{\partial x_1^*}{\partial \pi_n} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{\partial x_n^*}{\partial p} & -\frac{\partial x_n^*}{\partial \pi_1} & \cdots & -\frac{\partial x_n^*}{\partial \pi_n} \end{pmatrix}$$

es simétrica y semidefinida positiva. Por tanto, se verifica que los elementos de su diagonal principal no pueden ser negativos

$$\frac{\partial q^*}{\partial p} \geq 0, \quad \frac{\partial x_i^*}{\partial \pi_i} \leq 0 \quad \text{para cada } i = 1, \dots, n \quad (11.29)$$

es decir, un aumento en el precio del bien final conlleva una mayor oferta del mismo mientras que un aumento en el precio de un factor de producción genera una disminución de dicha oferta.

La minimización de los costes

En la sección anterior hemos analizado el comportamiento de una empresa cuyo objetivo era maximizar los beneficios. No obstante, hay en la teoría de la empresa otra cuestión de indudable interés económico, relacionada en cierta medida con la primera, pero independiente de ella. Se trata de la minimización de los costes de la empresa fijado un nivel mínimo de producción. Así, ahora estamos interesados en estudiar el problema

$$\left\{ \begin{array}{ll} \text{MIN.} & G(\pi, x) = \sum_{i=1}^n \pi_i x_i \\ \text{sujeto a} & q \leq f(x) \\ & x \geq 0 \end{array} \right. \quad (11.30)$$

conocidos $q > 0$ y $\pi > 0$.

Como antes, las hipótesis (H.1), (H.2), (H.3) y (H.4) garantizan existencia de única solución y permiten escribir (11.30) de la siguiente forma simplificada:

$$\left\{ \begin{array}{ll} \text{MIN.} & G(\pi, x) \\ \text{sujeto a} & f(x) = q \end{array} \right. \quad (11.31)$$

Los puntos estacionarios del programa (11.31) se obtienen aplicando las condiciones necesarias de primer orden de Lagrange. En concreto, si consideramos la función Lagrangiana

$$L(\pi, x, \lambda) = G(\pi, x) - \lambda \cdot (f(x) - q)$$

dichos puntos verifican el sistema

$$\left\{ \begin{array}{l} \frac{\partial L}{\partial x_i}(\pi, x, \lambda) = \pi_i - \lambda \cdot \frac{\partial f}{\partial x_i} = 0 \quad (i = 1, \dots, n) \\ \frac{\partial L}{\partial \lambda}(\pi, x, \lambda) = -f(x) + q = 0 \end{array} \right. \quad (11.32)$$

de donde se deduce que

$$\begin{cases} \frac{\partial f}{\partial x_i} = \frac{\partial f}{\partial x_j} \\ \pi_i = \pi_j \\ f(x) = q \end{cases} \quad (11.33)$$

siendo el multiplicador de Lagrange asociado

$$\lambda = \frac{\pi_i}{\frac{\partial f}{\partial x_i}(x)} > 0 \quad (11.34)$$

Como se observa, este problema es completamente análogo al correspondiente para la minimización del gasto en la teoría del consumidor; basta cambiar el papel de la utilidad por el de la producción. Del mismo modo, el sistema (11.32) genera las siguientes funciones de clase C^1 : la función solución $x^c(\pi, q) : \mathbb{R}_{++}^{n+1} \rightarrow \mathbb{R}^n$, denominada *función de demanda condicional*; la función valor óptimo $C(\pi, q) = G[x^c(\pi, q)] : \mathbb{R}_{++}^{n+1} \rightarrow \mathbb{R}$, denominada *función de costes de la empresa*, que indica el nivel de costes así alcanzado; y la *función multiplicador* $\lambda^c(\pi, q) : \mathbb{R}_{++}^{n+1} \rightarrow \mathbb{R}$, dada por (11.34).

Las propiedades de estas funciones son las mismas que las obtenidas en la minimización del gasto:

- (i) La función de demanda condicional $x^c(\pi, q)$ es homogénea de grado cero en la variable π , para cada q fijado.
- (ii) La función de costes $C(\pi, q)$ es homogénea de grado uno en la variable π , para cada q fijado.
- (iii) $x_i^c(\pi, q) = \frac{\partial C}{\partial \pi_i}$ para cada $i = 1, \dots, n$ (*lema de Shephard*).
- (iv) La función de costes $C(\pi, q)$ es cóncava en la variable π , para cada q fijado, y de clase C^2 .

Como consecuencia, la matriz hessiana $D_\pi^2 C(\pi, q)$ (denominada *matriz de Slutsky de la empresa*) es simétrica, semidefinida negativa y viene dada por

$$S(\pi, q) = \begin{pmatrix} \frac{\partial x_1^c}{\partial \pi_1} & \dots & \frac{\partial x_1^c}{\partial \pi_n} \\ \vdots & & \vdots \\ \frac{\partial x_n^c}{\partial \pi_1} & \dots & \frac{\partial x_n^c}{\partial \pi_n} \end{pmatrix}$$

Por tanto, se verifica que los elementos de su diagonal principal no pueden ser positivos

$$\frac{\partial x_i^c}{\partial \pi_i} \leq 0 \quad \text{para cada } i = 1, \dots, n \quad (11.35)$$

es decir, el incremento del precio de un input produce una disminución en su demanda condicional.

Si ahora fijamos los precios de los inputs resulta la curva de costes

$$C(q) \equiv C(\pi, q)$$

que relaciona costes y producción. Su derivada verifica

$$C'(q) = \frac{\partial C}{\partial q}(\pi, q) = \lambda^c(\pi, q) > 0 \quad (11.36)$$

En otras palabras, el multiplicador de Lagrange $\lambda^c(\pi, q)$ determina la función de coste marginal y la curva de costes es estrictamente creciente. Además, se puede comprobar que la curva de costes es también una función convexa: en efecto, sean q_1 y q_2 dos niveles de producción con vectores de inputs asociados (que resuelven los correspondientes problemas de minimización (11.30)) x^1 y x^2 , esto es

$$C(q_1) = \pi \cdot x^1, \quad C(q_2) = \pi \cdot x^2$$

Dado $\lambda \in [0, 1]$, $C(\lambda q_1 + (1 - \lambda)q_2)$ es el coste mínimo para producir un output $\lambda q_1 + (1 - \lambda)q_2$ y, al ser f cóncava

$$\lambda q_1 + (1 - \lambda)q_2 = \lambda f(x^1) + (1 - \lambda)f(x^2) \leq f(\lambda x^1 + (1 - \lambda)x^2)$$

por lo que $\lambda x^1 + (1 - \lambda)x^2$ es un vector de inputs factible en el problema de minimización de costes asociado a $q = \lambda q_1 + (1 - \lambda)q_2$. En consecuencia,

$$\begin{aligned} C(q) &= C(\lambda q_1 + (1 - \lambda)q_2) \leq G(\pi, \lambda x^1 + (1 - \lambda)x^2) = \\ &= \pi \cdot (\lambda x^1 + (1 - \lambda)x^2) = \lambda C(q_1) + (1 - \lambda)C(q_2) \end{aligned}$$

y la función $C(q)$ es convexa.

Para finalizar, vamos a analizar las relaciones existentes entre los problemas de maximización de los beneficios y minimización del coste. Aunque estos problemas no son exactamente equivalentes (duales), sí tienen una estrecha relación. En concreto, se verifican las siguientes propiedades (su comprobación se deja como ejercicio para el lector):

- (i) $x^*(p, \pi)$ es la única solución correspondiente a maximizar la función $\bar{B}(p, \pi, x) = p \cdot f(x) - C(\pi, f(x))$.
- (ii) $x^*(p, \pi) = x^c(\pi, q^*(p, \pi))$

Bibliografía

- [1] Barbolla, R., Cerdá, E. y Sanz, P. *Optimización. Cuestiones, ejercicios y aplicaciones en la economía*. Prentice Hall, Madrid, 2000
- [2] Barbolla, R. y Sanz, P. *Álgebra lineal y teoría de matrices*. Prentice Hall, Madrid, 1998
- [3] Cartan, H. *Cálculo diferencial*. Omega, Barcelona, 2ª edición, 1978
- [4] Cartan, H. *Formas diferenciales: aplicaciones elementales al cálculo de variaciones y a la teoría de curvas y superficies*. Omega, Barcelona, 1971
- [5] Dieudonné, J. A. *Fundamentos de análisis moderno*. Reverté, Barcelona, 1979
- [6] Fernández Pérez, C. y Vegas, J. M. *Ecuaciones diferenciales II. Ecuaciones no lineales*. Pirámide, Madrid, 1996
- [7] Gravelle, H. y Rees, R. *Microeconomía*. Alianza Universidad Textos, Madrid, 1984
- [8] Hernández, E. *Álgebra y Geometría*. Addison-Wesley Iberoamericana, Madrid, 1994
- [9] Madden, P. *Concauidad y optimización en microeconomía*. Alianza Universidad, Madrid, 1987
- [10] Rubio, B. y Guzmán, M. *Problemas, conceptos y métodos del análisis matemático*. 3 volúmenes; Vol. 1: Números reales, sucesiones y series; Vol. 2: Funciones, integrales, derivadas; Vol. 3: Sucesiones y series de funciones, números complejos. Pirámide, Madrid, 1990-92-93
- [11] Samuelson, P. A. y Nordhaus, W. D. *Economía*. McGraw-Hill, Madrid, 16ª edición, 1999

Índice de Materias

Aceleración

escalar, 441

vector, 441

Adherencia de un conjunto, 70, 368

Agentes económicos, 483, 499

intercambio entre, 496

Aplicación contractiva, 384, 454

Aproximaciones sucesivas

método, 454

Arco, 396

Base, 9

canónica, 9

Beneficio

de la empresa

función de, 521

función de, 488, 520

maximización del, 520

Bien(es)

de lujo, 511

Giffen, 518

inferior, 510

necesario, 511

normal, 511

precio de equilibrio de un, 487

precio de un, 484

complementarios perfectos, 504

complementarios hicksianos netos, 516

sustitutivos perfectos, 503

sustitutivos hicksianos netos, 516

valor de un, 484

Bola

abierta, 61, 367, 390, 404

cerrada, 61, 369

Caja de Edgeworth, 498

Cálculo aproximado, 113

Cónicas, 16

degeneradas, 20

Cambio de variables, 140, 181

Campo vectorial, 215, 442

gradiente, 112, 215, 444

Cauchy

criterio para la convergencia uniforme de, 386

sucesión de, 372, 405

Cicloide, 436

Cilindro, 36, 40

Clase C^p , 211, 431

Cociente incremental parcial, 93

Componente conexa, 422

Composición de funciones, 79, 136, 381, 428

Condición

de Lipschitz, 384

de regularidad, 317, 322, 346, 350

Condiciones

de Kuhn-Tucker, 344, 346, 475

de Lagrange, 315, 322, 473

de primer orden, 276, 301

de segundo orden, 282, 286, 305, 324, 352

Conjunto

- abierto, 68, 367, 402
 - relativo, 388
- acotado, 67, 371, 405
- adherencia de un, 70, 368
- cerrado, 68, 368, 402
 - relativo, 388
- compacto, 374, 403
- conexo, 151, 395, 403
 - por arcos, 397
 - por poligonales, 151
- convexo, 12, 59
- de indiferencia, 500
- de nivel, 42, 61
- denso, 412
- diámetro de un, 371
- exterior de un, 368
- factible, 48, 310, 320, 340
- frontera de un, 68, 368
- interior de un, 68, 368

Cono, 40, 155

Continuidad

- de una función, 65, 78, 379, 380
- uniforme, 78, 383

Convergencia

- absoluta de una serie, 374
- de una serie, 374
- de una serie de funciones, 387
- de una sucesión, 370
- de una sucesión de funciones, 385
- uniforme, 386, 387

Coordenadas

- cartesianas, 8, 24
- cilíndricas, 41
- esféricas, 41
- polares, 22

Coste

- de la empresa
 - función de, 523
- función de, 488, 493

- de oportunidad, 484
- de producción, 484
- marginal, 488
- minimización del, 522

Cuádriga, 39

Curva, 435

- de contrato, 498
- de costes, 524
- de Engel, 511
- de indiferencia, 500, 502
- de nivel, 42, 169
- diferenciable, 439
- integral de un campo vectorial, 442
- marshalliana, 491
- orientación en una, 438
- recta tangente a una, 440

Demanda

- condicional
 - función de, 523
- de factores
 - función de, 521
- elástica, 490, 492
- elasticidad de la, 489
- función de, 486
- función hicksiana de, 513
- función marshalliana de, 508
- inelástica, 489, 492
- marginal, 478
- teoría clásica de la, 500

Dependencia funcional, 185, 470

Derivada(s)

- de la función implícita, 165, 172, 175, 179, 467
- de la función inversa, 183, 470
- de orden superior, 211, 221
- direccional, 143, 149
 - fórmula de la, 146, 149
 - máxima, 148
- parcial(es), 93, 98

- interpretación geométrica de la, 95
- según un vector, 196, 431, 449
- según una curva, 197
- según una recta, 196
- segundas, 212
- y continuidad, 97
- Desigualdad
 - de Cauchy-Schwarz, 15, 56, 408
 - de Hölder, 407, 415
 - de Minkowski, 407, 416
 - triangular, 16, 57, 367, 404, 406
- Determinante jacobiano, 178, 463
- Diferenciabilidad y continuidad, 109
- Diferencial, 90, 103, 117, 424
 - de Fréchet, 448
 - de Gateaux, 449
 - de orden superior, 248, 445
 - interpretación geométrica de la, 110
 - parcial, 459
 - segunda, 233, 238, 445
- Distancia, 13, 28, 58, 366, 390, 404, 406
 - entre dos conjuntos, 413
- Dominio de una función, 4
- Dualidad, 495, 512
 - relaciones de, 516
- Ecuación
 - de Laplace, 268
 - de ondas, 268
 - de Slutsky, 517
 - del calor, 268
 - diferencial ordinaria, 206, 443
 - en derivadas parciales, 133, 268
- Efecto
 - renta, 518
 - sustitución, 518
- Elasticidad, 488
 - constante
 - función de, 491
 - de la demanda, 489
 - parcial, 491
 - y transformación logarítmica, 490
- Elipse, 16
- Elipsoide, 37
- Energía
 - cinética, 451
 - potencial, 451
- Engel
 - curva de, 511
 - ley de, 512
- Entorno, 67, 367, 402
 - relativo, 388
 - ε -entorno, 67, 367
- Esfera, 35, 61
- Espacio
 - de Banach, 407
 - de Hilbert, 408
 - euclídeo, 408
 - métrico, 58, 404
 - completo, 405
 - normado, 58, 405
 - topológico, 402
 - vectorial, 54
- Extensión de una función, 382, 415
- Forma bilineal, 447
- Forma cuadrática, 284, 447
 - clasificación de una, 285, 287, 290, 293, 294, 337
 - definida positiva, 285
 - definida negativa, 285
 - indefinida, 286
 - semidefinida positiva, 285
 - semidefinida negativa, 285
- Fórmula
 - de los incrementos finitos, 150, 434
 - del incremento, 90, 101, 102
 - de Taylor, 231, 242

Función

- cóncava, 297, 298
- continua, 65, 78, 79, 380, 402
 - en un conjunto, 79, 380, 402
 - uniformemente, 79, 383
- continuamente diferenciable, 104, 122, 427
- convexa, 297, 298
- de beneficio, 488, 520, 521
- de clase C^1 , 104, 122, 427
- de clase C^2 , 214
- de clase C^p , 221, 431
- de Cobb-Douglas, 50, 491
- de coste, 488, 493, 523
- de demanda, 484
 - de factores, 521
 - condicional, 523
 - hicksiana, 513
 - marshalliana, 508
- de gasto, 505, 514
- de oferta, 486, 521
- de producción, 3, 339, 486
- de utilidad, 478, 498, 500
 - indirecta, 508
- diferenciable, 103, 119, 423
 - en un conjunto, 107, 423
- escalar, 4
- dominio de una, 4
- gráfica de una, 5, 34
- homogénea, 155
- imagen de una, 5
- imagen inversa por una, 387
- implícita, 21, 162
- independiente de una variable, 153, 155
- inversa, 181
- lagrangiana, 315
- límite de una, 71, 377
- lineal, 6
- lipschitziana, 155, 384
- objetivo, 48, 310

- parcial, 92
- potencial, 215
- rango de una, 5
- recorrido de una, 5
- vectorial, 4, 365

Frontera, 68, 368

Gasto

- del consumidor
 - función de, 514
- función de, 505
- minimización del, 512

Giffen

- bienes, 518
- paradoja de, 517

Gradiente, 34, 61, 112, 118

Gráfica, 5, 34

Hélice, 436

Hipérbola, 16, 18

Hiperboloide, 37

- de una hoja, 37
- de dos hojas, 38

Hiperplano, 7, 59

- tangente, 119

Hipersuperficie, 62

Hessiana, matriz, 238, 241, 448

Hessiano, 238, 241

Hessiano orlado, 326, 328, 331

Identidad de Roy, 509

Imagen, 5

- de un conjunto compacto, 382
- de un conjunto conexo, 396
- inversa, 387
 - de un conjunto abierto, 388
 - de un conjunto cerrado, 388

Incrementos parciales, 93

Indiferencia

- conjuntos de, 500
- curvas de, 500, 502

Inf, ínfimo, 46

- Infinitésimo, 90, 105, 120
- Interior de un conjunto, 68
- Isocuantas, 519
- Jacobiano, 178, 463
- Kuhn-Tucker
 - multiplicadores de, 347, 354
 - teorema de, 344, 346, 475
- Lagrange
 - multiplicadores de, 315, 322
 - teorema de, 322, 473
- Lagrangiano, 315, 322, 344
- Lema
 - de Hotelling, 521
 - de Shepard, 515, 523
- Ley(es)
 - de Engel, 512
 - de Morgan, 411
 - del paralelogramo, 8, 409
- Ligadura, 312
- Límite
 - de una función, 71, 377
 - de una sucesión, 370
 - direccional, 379
- Longitud de un vector, 13, 28, 55
- Matriz
 - de Slutsky, 515, 523
 - exponencial de una, 419
 - jacobiana, 183
 - hessiana, 238, 241, 448
- Máximo, 43, 282
 - global, 275, 282
 - libre, 311
 - local, 282
 - restringido, 312
- Método
 - de la cuerda, 455
 - de mínimos cuadrados, 307
 - de Newton, 457
- Métrica, 58, 404
- Mínimo, 43, 282
 - global, 275, 282
 - local, 282
- Módulo de un vector, 13, 28, 55
- Multíndice, 244
- Multiplicadores de Lagrange, 315, 322
 - significado de los, 332
- Multiplicadores de Kuhn-Tucker, 347, 354
 - significado de los, 353
- Norma(s)
 - de un vector, 13, 28, 55, 58, 366, 390, 405
 - de una aplicación lineal, 393
 - equivalentes, 391
- Oferta
 - del bien
 - función de, 521
 - función de, 486
 - marginal, 488
 - teoría clásica de la, 518
- Optimización, 42, 275
 - con restricciones de igualdad, 309
 - con restricciones de desigualdad, 339
 - problemas convexos de, 297, 306, 323, 351
- Óptimo, 275
 - de Pareto, 498
 - local, 282
 - global, 282
- Parábola, 16, 18
- Paraboloide, 38
- Pareto
 - eficiente en sentido de, 498
 - óptimo de, 498

- Plano(s), 6, 25
 ecuación(es)
 explícita, 28
 implícita, 27
 paramétricas, 27
 vectorial, 26
 normal a una curva, 450
 ortogonales, 30
 pendiente de un, 32, 34
 tangente, 111
 variedad de dirección de un, 26
- Poligonal, 151
- Producción
 coste de, 484
 función de, 486, 518
 Cobb-Douglas, 50, 491
- Producto escalar, 14, 29, 56
- Programa lineal, 46
- Puerto, 283
- Punto, 53
 aislado, 71, 370
 crítico, 170, 277
 de acumulación, 71, 370, 402
 de adherencia, 370
 de equilibrio, 444
 de silla, 283
 estacionario, 316
 exterior, 68, 368, 402
 frontera, 68, 368, 402
 interior, 68, 368, 402
- Rango, 5
- Recorrido, 5
- Recta(s), 10, 59
 de regresión, 309
 ecuación(es)
 en forma continua, 25
 implícita, 11
 explícita, 11
 paramétricas, 11, 25
 vectorial, 11, 24
 ortogonales, 14, 59
 paralelas, 10
 pendiente de una, 12
 soporte, 44
 tangente, 440
- Recubrimiento, 375
 abierto, 375
- Regla
 de la cadena, 136, 428
 de Leibniz, 157
- Relación marginal
 de sustitución, 501
 de sustitución técnica, 519
- Restricción(es)
 de desigualdad, 47, 339
 de igualdad, 309
 de una función, 381
 saturada, 342
- Resto (o término complementario),
 90, 101, 103, 106, 233
 de Lagrange, 234, 241, 243, 248,
 336
 de Peano, 238, 241, 248
 distintas expresiones del, 259
 expresión equivalente del, 108
- \mathbb{R}^n , 4, 52
- Segmento de recta, 12, 59
- Semiespacio, 60
- Semiplano, 44
- Serie, 373
 convergencia absoluta de una,
 374
 convergencia de una, 374
 criterio de comparación, 374
 de funciones, 387
 convergencia uniforme de una,
 387
- Silla de montar, 38
- Sistema gradiente, 450
- Slutsky
 ecuación de, 517
 matriz de, 515, 523

Solución factible, 48, 310, 320, 340

Subespacio vectorial, 55

Sucesión

acotada, 372

convergente, 370

de Cauchy, 372, 405

de funciones, 385

límite de una, 370, 403

Sup, supremo, 46

Teorema

de Banach, 454

de Bolzano, 396

de Bolzano-Weierstrass, 372, 373

de Euler, 156

de Heine-Borel, 375

de la aplicación contractiva, 454

de la contracción uniforme, 458

de la envolvente, 337

de la función implícita, 165, 175,
178, 462

de la función inversa, 181, 469

de Lagrange, 322, 473

de Kuhn-Tucker, 343, 346, 475

de Pitágoras, 13, 28, 55

de Schwarz, 256, 448

de Weierstrass, 276, 382

del punto fijo, 420, 454

del valor medio, 150, 154, 432

Término complementario (véase Resto)

Topología, 402

de \mathbb{R}^n , 389

relativa, 388, 402

Trayectoria, 438, 443

Unicidad

de la diferencial, 105

del polinomio de Taylor, 252

Utilidad, 484

función de, 486, 498, 500

indirecta

función de, 508

marginal, 485, 488, 498

maximización de la, 505

Variables de decisión, 309

Vector(es), 8, 24, 53

combinación lineal de, 9, 54

dirección, 11

longitud de un, 13, 28, 55

módulo, 13

multiplicación por escalares, 9,
24, 54

norma, 13, 28, 55, 58, 366, 390,
405

normal, 29

ortogonales, 13, 29

suma de, 8, 24, 54

tangente a una curva, 440

Velocidad

media, 441

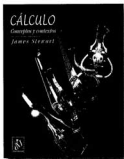
instantánea, 441

vector, 441

Otras obras afines

publicadas por **THOMSON**

Cálculo. Conceptos y contextos



ISBN: 968-7529-60-1

James Stewart

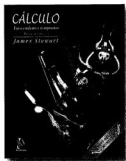
El resultado de la reforma en la enseñanza del cálculo es la regla de tres, la cual establece que los temas deben presentarse de manera geométrica, numérica y algebraica. El autor propone en el presente libro un cuarto criterio para completar el proceso del aprendizaje: la parte verbal o descriptiva. Lo anterior es un aspecto del nuevo enfoque que ahora nos presenta Stewart: el énfasis en la transmisión de los "conceptos". Esta obra está escrita para llevarse como libro, de texto en un curso de un año o durante dos semestres.

Cálculo. Trascendentes tempranas

ISBN: 968-7529-92-X

James Stewart

Un libro más del autor de cálculo más reconocido a nivel mundial. Siguiendo la misma línea de sus otros libros, en esta obra se hace énfasis en la comprensión, presenta detalle matemático suficiente para la descripción precisa, sin permitir que el formalismo se tome engorroso. El espíritu de reforma está presente en el libro, pero dentro del contexto de los requisitos tradicionales.



Cálculo Diferencial de Varias Variables

En la primera parte, este texto presenta el cálculo diferencial de varias variables de la forma más intuitiva posible. Se ha procurado una redacción flexible, dejando para el final de cada apartado los conceptos más avanzados y las demostraciones más complicadas, a fin de que la presentación de la materia y sus aplicaciones más comunes e importantes resulte accesible, fluida y sin interrupciones.

La segunda parte contiene temas de carácter más avanzado y en ella se justifican con rigor los -escasos- resultados difíciles que se han utilizado sin demostración en la primera parte. Estos temas no son indispensables para un estudiante de los primeros años de ciencias sociales o biológicas, pero satisfacen ampliamente las necesidades de los cursos superiores de índole más cuantitativa de dichas carreras. Por otro lado, cubren, sin duda, el nivel con que se imparte la disciplina en las facultades científicas y técnicas.

La tercera parte contiene un breve resumen de las aplicaciones más típicas del cálculo diferencial a la economía. Está dirigida a estudiantes de Económicas y Empresariales, y también a los de otras titulaciones científicas y técnicas cuyos planes de estudios más recientes incluyen asignaturas de Economía. Esperamos que suponga también un estímulo para estudiantes de otras carreras, que pueden hallar en estos capítulos una introducción elemental a aplicaciones de las matemáticas que quizá les sorprendan.

